

# BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

**COPIE OFFICIELLE**

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 29 MARS 2002

Pour le Directeur général de l'Institut  
national de la propriété industrielle  
Le Chef du Département des brevets

Martine PLANCHE

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

**INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE**

**SIEGE**  
26 bis, rue de Saint Petersburg  
75800 PARIS cedex 08  
Téléphone : 33 (1) 53 04 53 04  
Télécopie : 33 (1) 42 93 59 30  
[www.inpi.fr](http://www.inpi.fr)

ETABLISSEMENT PUBLIC NATIONAL CREE PAR LA LOI N° 51-444 DU 19 AVRIL 1951

**BEST AVAILABLE COPY**



INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INTELLECTUELLE

26 bis, rue de Saint Pétersbourg

75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

# BREVET D'INVENTION CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI



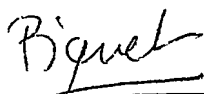
N° 11354\*01

REQUÊTE EN DÉLIVRANCE 1/2

Cet imprimé est à remplir lisiblement à l'encre noire

DE 510 W / 260899

<b>REMISE EN DÉPÔT</b> DATE <b>16 MARS 2001</b> LIEU <b>75 INPI PARIS</b> N° D'ENREGISTREMENT <b>0103631</b> NATIONAL ATTRIBUÉ PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI <b>16 MARS 2001</b>		<b>11 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE</b> ERNEST GUTMANN-YVES PLASSERAUD S.A. 3 rue Chauveau-Lagarde 75008 PARIS	
<b>Vos références pour ce dossier (facultatif)</b> B4787-FL			
<b>Confirmation d'un dépôt par télécopie</b> <input type="checkbox"/> N° attribué par l'INPI à la télécopie			
<b>2 NATURE DE LA DEMANDE</b>		<b>Cochez l'une des 4 cases suivantes</b>	
Demande de brevet		<input checked="" type="checkbox"/>	
Demande de certificat d'utilité		<input type="checkbox"/>	
Demande divisionnaire		<input type="checkbox"/>	
Demande de brevet initiale		N° _____ Date ____/____/____	
ou demande de certificat d'utilité initiale		N° _____ Date ____/____/____	
Transformation d'une demande de brevet européen		N° _____ Date ____/____/____	
<b>3 TITRE DE L'INVENTION (200 caractères ou espaces maximum)</b>			
MOLECULES D'ACIDES NUCLEIQUES CODANT UNE DEXTRANE- SACCHARASE CATALYSANT LA SYNTHÈSE DE DEXTRANE PORTANT DES RAMIFICATIONS DE TYPE ALPHA-1,2 OSIDIQUES.			
<b>4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE</b>		Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»	
<b>5 DEMANDEUR</b>		<input checked="" type="checkbox"/> S'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé «Suite»	
Nom ou dénomination sociale		INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE TOULOUSE	
Prénoms			
Forme juridique		Etablissement public national à caractère scientifique, culturel et professionnel	
N° SIREN			
Code APE-NAF			
Adresse		Rue 135 avenue de Rangueil	
		Code postal et ville 31077 TOULOUSE CEDEX 4	
Pays		FRANCE	
Nationalité		Française	
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			

REMISE DES PIÈCES DATE <b>18 MARS 2001</b> LIEU <b>75 INPI PARIS</b> N° D'ENREGISTREMENT <b>0103631</b> NATIONAL ATTRIBUÉ PAR L'INPI		Réservé à l'INPI
<b>Vos références pour ce dossier :</b> <i>(facultatif)</i>		B4787-FL
<b>6 MANDATAIRE</b>		
Nom		LAZARD
Prénom		Florence
Cabinet ou Société		ERNEST GUTMANN-YYES PLASSERAUD S.A.
N° de pouvoir permanent et/ou de lien contractuel		
Adresse	Rue	3 rue Chauveau-Lagarde
	Code postal et ville	75008 PARIS
N° de téléphone <i>(facultatif)</i>		01 44 51 18 00
N° de télécopie <i>(facultatif)</i>		01 42 66 08 90
Adresse électronique <i>(facultatif)</i>		info@egyp.fr
<b>7 INVENTEUR (S)</b>		
Les inventeurs sont les demandeurs		<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée
<b>8 RAPPORT DE RECHERCHE</b>		Uniquement pour une demande de brevet (y compris division et transformation)
Établissement immédiat ou établissement différé		<input checked="" type="checkbox"/> <input type="checkbox"/>
Paiement échelonné de la redevance		Paiement en trois versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input type="checkbox"/> Non
<b>9 RÉDUCTION DU TAUX DES REDEVANCES</b>		Uniquement pour les personnes physiques <input type="checkbox"/> Requête pour la première fois pour cette invention (joindre un avis de non-imposition) <input type="checkbox"/> Requête antérieurement à ce dépôt (joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence) :
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes		1
<b>10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE</b> (Nom et qualité du signataire) LAZARD Florence CPI n° 92-4029		VISA DE LA PRÉFECTURE OU DE L'INPI 

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.



INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

# BREVET D'INVENTION

## CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI



N° 11354\*01

### REQUÊTE EN DÉLIVRANCE

Page suite N° 1.../1...

REMISSÉ PAR	16 MARS 2001
DATE	75 INPI PARIS
LIEU	
N° D'ENREGISTREMENT	0103631
NATIONAL ATTRIBUÉ PAR L'INPI	

Cet imprimé est à remplir lisiblement à l'encre noire

DS 829 W / 260899

Vos références pour ce dossier (facultatif)		B4787-FL	
<input checked="" type="checkbox"/> DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE		Pays ou organisation	
		Date / / N°	
		Pays ou organisation	
		Date / / N°	
		Pays ou organisation	
		Date / / N°	
<input checked="" type="checkbox"/> DEMANDEUR			
Nom ou dénomination sociale		CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	
Prénoms			
Forme juridique		Etablissement public	
N° SIREN			
Code APE-NAF			
Adresse	Rue	3 rue Michel-Ange	
	Code postal et ville	75794	PARIS
Pays		FRANCE	
Nationalité		Française	
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			
<input checked="" type="checkbox"/> DEMANDEUR			
Nom ou dénomination sociale			
Prénoms			
Forme juridique			
N° SIREN			
Code APE-NAF			
Adresse	Rue		
	Code postal et ville		
Pays			
Nationalité			
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			
<input checked="" type="checkbox"/> SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire) LAZARD Florence CPI n° 92-4029		VISA DE LA PRÉFECTURE OU DE L'INPI Riquel	

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire.  
Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI

**MOLECULES D'ACIDES NUCLEIQUES CODANT UNE DEXTRANE-  
SACCHARASE CATALYSANT LA SYNTHÈSE DE DEXTRANE  
PORTANT DES RAMIFICATIONS DE TYPE ALPHA-1,2 OSIDIQUES**

5

La présente invention relève du domaine de la glycotechnologie et plus particulièrement de la synthèse d'oligosaccharides ou oligosides à effet prébiotique, thérapeutique ou diagnostique.

10

La présente invention porte sur des molécules d'acides nucléiques codant une enzyme ayant une activité de glycosyltransférase catalysant la synthèse de dextranes ou d'oligosides portant des ramifications de type  $\alpha(1 \rightarrow 2)$  osidiques.

15

L'invention porte en outre sur les enzymes synthétisées par les acides nucléiques selon l'invention, ainsi que sur leurs systèmes d'expression dans des cellules procaryotes ou eucaryotes. Elles portent enfin sur l'utilisation desdites enzymes dans la production d'oligosaccharides dans l'alimentation, ou en tant que principe actif de produits thérapeutiques et/ou cosmétiques.

20

Les oligosides et hétérooligosides jouent le rôle de signaux de reconnaissance et d'effecteur chez l'animal comme dans les plantes (on parle alors d'oligosaccharines), en se liant spécifiquement à des lectines, des glycosyltransférases, des glycosidases, des molécules d'adhésion, etc... Ainsi, les déterminants antigéniques des groupes sanguins sont des osides, et notre défense contre nombre de bactéries pathogènes est dirigée contre les structures osidiques de l'enveloppe bactérienne. Par ailleurs, l'une des raisons majeures du rejet des xénogreffes est l'existence de structures osidiques propres à chaque espèce. Ces propriétés, ainsi que les connaissances acquises ces dernières années sur les glycosyltransférases et les lectines, contribuent à faire de certains oligosides des candidats de choix pour la thérapeutique ou la prophylaxie des désordres liés à l'équilibre microbiologique de différents organes tels

25

30

l'intestin, ou la peau. Par exemple, les oligosides constituent une alternative intéressante à l'utilisation de microorganismes et d'antibiotiques pour réguler la composition de la flore intestinale (effet prébiotique). Certains oligosides peuvent être considérés comme des "fibres solubles" lorsqu'ils ne sont pas métabolisés par les enzymes digestives humaines et animales ; en gagnant le colon, ils interagissent avec la flore microbienne et affectent spécifiquement la croissance et l'adhésion de certaines espèces. Incorporées à faible dose (moins de 1 %) dans l'alimentation, certaines de ces molécules osidiques améliorent l'état de santé et stimulent la prise de poids des animaux.

Une revue des différentes glycosyltransférases, leur structure et leur activité, est décrite dans Vincent Monchois et al. (réf. 1). Brièvement :

a) Il apparaît que la structure des glycosyltransférases et/ou dextrane-saccharases étudiées est très conservée et est constituée, partant de la partie aminée de la protéine, d'une séquence signal, d'un domaine variable, d'un domaine catalytique et d'un domaine de liaison au glucane.

b) Les glucooligosides (GOS) sont synthétisables par des glycosyltransférases telles les dextrane-saccharases, à partir de substrats peu coûteux tel le saccharose et en présence d'un sucre accepteur de glucose. D'autres substrats, tels l' $\alpha$ -D-fluoro-glucose, le paranitrophényl- $\alpha$ -D-glucopyranoside, l' $\alpha$ -D-glucopyranoside- $\alpha$ -D-sorbofuranoside ou le 4-O- $\alpha$ -D-galactopyranosylsucrose peuvent également être utilisés.

Ces enzymes catalysent à partir du substrat le transfert d'unités glucose sur des molécules acceptrices. En présence d'un accepteur de glucose tel le maltose, ou l'isomaltose, les glycosyltransférases catalysent la synthèse d'oligosaccharides de bas poids moléculaire comprenant majoritairement des chaînes de 3 à 7 glucoses. En revanche, en absence d'accepteur, l'enzyme synthétise des glucanes de haut poids moléculaire de type dextrane.

c) Les structures et la fonction des glucanes ou des oligosides synthétisés par les glycosyltransférases dépendent de la souche bactérienne productrice.

Dans l'ensemble de ce texte, on appellera de façon générique des glycosyltransférases les différentes enzymes capables de catalyser la synthèse de polymères de glucose à partir de saccharose. Elles sont généralement produites par des souches bactériennes de type *Leuconostoc*, *Lactococcus*, *Streptococcus* ou *Neisseria*. La taille et la structure des glucanes produits dépendent de la souche productrice.

Les unités de glucose sont couplées par des liaisons osidiques  $\alpha(1\rightarrow6)$  comme dans le dextrane, par des liaisons  $\alpha(1\rightarrow3)$ , comme dans le cas du mutane, ou par une alternance des deux types (alternane).

De la même façon, l'existence et la nature des ramifications, leur longueur et leur position varient selon l'origine de la souche productrice.

Les glycosyltransférases produisant des glucanes ou des GOS contenant au moins 50 % de liaison  $\alpha(1\rightarrow6)$  sont appelées dextrane-saccharases. Celles ci sont produites notamment par des bactéries de type *Leuconostoc mesenteroides*.

d) La dextrane-saccharase de *L. mesenteroides* NRRL B-1299 a la particularité de produire, quant à elle, un dextrane hautement ramifié dont la majorité des ramifications sont de type  $\alpha(1\rightarrow2)$ . Utilisée en présence de saccharose et de maltose, molécule acceptrice de glucose, elle conduit à la formation de GOS présentant pour certains une liaison  $\alpha(1\rightarrow2)$  à leur extrémité non réductrice et pour d'autres des ramifications  $\alpha(1\rightarrow2)$  sur les résidus intermédiaires entre les extrémités. A ce titre, ils résistent à la dégradation par les enzymes (hydrolases) du tractus digestif supérieur, chez l'homme et l'animal, et ne sont dégradés que par les bactéroïdes, bénéfiques à l'organisme. Un phénomène identique se produit au niveau de la peau, permettant d'envisager des applications en

cosmétologie, car c'est le déséquilibre de la flore microbienne cutanée qui est à l'origine de nombreux problèmes cosmétiques et dermatologiques. C'est en raison de ces caractéristiques qu'ils sont désignés ici par le terme GOS d'intérêt.

5 Dans l'ensemble du texte, les polysaccharides synthétisés par les glycosyltransférases selon l'invention sont soit des dextrans de haut poids moléculaire lorsque la réaction est réalisée sans accepteur de glucose, soit des oligosides lorsque la réaction est réalisée en présence d'accepteur de glucose tel le maltose ou l'isomaltose sans que cela soit  
10 nécessairement spécifié. En effet, la fonctionnalité de l'enzyme est caractérisée par la nature des liaisons glucose-glucose ( $\alpha(1\rightarrow6)$ ,  $\alpha(1\rightarrow2)$ ) ou autres et non par le poids moléculaire du polysaccharide synthétisé.

15 Les dextrane-saccharases de *L. mesenteroides* trouvent déjà de nombreuses applications dans l'industrie, et en particulier celles de la souche NRRL B-1299 pour lesquelles un procédé de synthèse des GOS présentant des ramifications  $\alpha(1\rightarrow2)$  a été décrit dans le brevet EP 0325 872 B1.

20 Marguerite Dols et al. (2) ont montré que les GOS produits par les dextrane-saccharases de cette souche sont en fait un mélange d'au moins trois familles de molécules similaires différant de fait par le nombre et le positionnement des ramifications de type  $\alpha(1\rightarrow2)$ , ce qui amène l'hypothèse de l'existence de différentes activités enzymatiques de type glycosyltransférase dans cette souche bactérienne.

25 Compte tenu de l'intérêt industriel dans le domaine des aliments prébiotiques, en cosmétologie ou en pharmacie des GOS présentant des ramifications  $\alpha(1\rightarrow2)$  et rappelé ci-dessus, la présente invention vise à isoler et caractériser une enzyme particulière parmi celles produites par *L. mesenteroides* NRRL B-1299 qui serait plus particulièrement impliquée dans la synthèse d'oligosides présentant les  
30 ramification  $\alpha(1\rightarrow2)$ . L'identification et la caractérisation d'une telle enzyme offrent l'avantage, d'une part, de fournir un procédé de production uniforme



et reproductible des GOS d'intérêt et, d'autre part, d'identifier les caractéristiques essentielles de l'enzyme productrice de ces GOS d'intérêt, afin, le cas échéant, d'améliorer les performances des produits de la réaction enzymatique en fonction de l'utilisation envisagée.

5 Le problème technique sous-tendu dans la présente invention était ainsi de pouvoir disposer d'une enzyme et donc des acides nucléiques isolés codant cette enzyme permettant la production améliorée de GOS à ramifications  $\alpha(1\rightarrow2)$ .

10 La présente invention apporte une solution technique aux différentes questions évoquées ci-avant en fournissant une nouvelle dextrane-saccharase, appelée DSR-D codée par un gène doté d'une structure nouvelle et inattendue (*dsrD*) et capable de catalyser la synthèse des glucanes ou des oligosaccharides contenant des ramifications  $\alpha(1\rightarrow2)$ . Par structure nouvelle et inattendue, on entend le fait que l'organisation de  
15 la protéine diffère de celle de toutes les autres glycosyltransférases décrites à ce jour (1) et dont le domaine catalytique est situé en amont d'un domaine de liaison au glucane, ce dernier constituant la partie carboxylique de la protéine.

20 Ainsi, la présente invention porte sur un polypeptide isolé ayant une activité enzymatique de glycosyltransférase apte à former des dextrans présentant des ramifications  $\alpha(1\rightarrow2)$ , caractérisé en ce qu'il comprend au moins un domaine de liaison au glucane et un domaine à activité catalytique situé en aval du domaine de liaison au glucane. Par  
25 situé en aval, on entend le fait que la partie aminée de la séquence à activité catalytique ou Domaine catalytique est proximale de la partie carboxylique du domaine de liaison au glucane. Ces deux domaines peuvent être immédiatement contigus ou au contraire séparés par une région variable.

30 La glycosyltransférase selon l'invention comporte de préférence un peptide signal.

Dans un mode de réalisation de l'invention, la glycosyltransférase comprend deux domaines catalytiques situés de part et d'autre du domaine de liaison au glucane.

La présence d'un domaine à activité catalytique dans la partie carboxylique de l'enzyme est une caractéristique essentielle de cette dernière dans sa capacité à former des liaisons  $\alpha(1\rightarrow2)$  osidiques. En effet, comme le montrent les expériences décrites ci-après, la délétion de ce domaine dans une enzyme ayant au moins deux domaines catalytiques conduit à la production de glucanes ou d'oligosides ayant essentiellement des liaisons osidiques de type  $\alpha(1\rightarrow6)$  et dépourvus de liaisons de type  $\alpha(1\rightarrow2)$ .

L'analyse comparative des différentes glycosyltransférases incluant les dextrane-saccharases a mis en évidence un très fort degré de conservation de leur domaine catalytique.

Le domaine catalytique situé dans la partie carboxy-terminale de la glycosyltransférase selon l'invention a une séquence présentant au moins 44 % d'identité et 55 % de similarité avec les domaines catalytiques des autres glycosyltransférases analysées. En particulier, le domaine catalytique dans la partie carboxylique de la glycosyltransférase selon l'invention a au moins 65 % d'identité et au moins 80 % de similarité avec la séquence ID No. 1 représentée dans la figure 7, la triade catalytique Asp/Glu/Asp en positions respectives 2210/2248/2322 étant conservée.

Dans l'ensemble du texte, on entend par X% de similarité par rapport à une séquence de référence le fait que X % des acides aminés sont identiques ou modifiés par substitution conservative telle que définie dans le logiciel d'alignement des séquences d'acides aminés ClustalW (<http://bioweb.pasteur.fr/docs/doc-gensoft/clustalw/>) et que (100-X) % peuvent être délétés, substitués par d'autres amino-acides, ou encore que (100-X) % peuvent être ajoutés à la séquence de référence.

Une structure primaire particulière de l'enzyme selon l'invention est représentée dans la séquence ID No. 2 qui représente une séquence de 2835 acides aminés d'une dextrane-saccharase de *L. mesenteroides* B1299.

5 Cette dextrane-saccharase, nommée DSR-D, possède comme la plupart des glycosyltransférases et des dextrane-saccharases une séquence signal, une région variable faiblement conservée, un domaine catalytique hautement conservé (CD1), un domaine de liaison au glucane (GBD) et un deuxième domaine catalytique (CD2) dans la partie  
10 carboxylique de la protéine. DSR-D est la première glycosyltransférase analysée et présentant deux domaines catalytiques, dans la configuration présentée dans la figure 1 b). C'est également la première glycosyltransférase dont un domaine catalytique est situé dans la partie carboxylique de la protéine.

15 La comparaison et l'analyse de la séquence de DSR-D avec les séquences des glycosyltransférases ou des dextrane-saccharases déjà décrites (1), ainsi que les moyens utilisés à cette fin sont indiqués dans l'exemple 2 détaillé ci-après. Il y apparaît clairement que si l'existence de deux domaines catalytiques différencie substantiellement DSR-D des  
20 autres enzymes, en revanche les séquences desdits domaines sont substantiellement conservées. En particulier, les acides aminés nécessaires à l'activité catalytique sont conservés dans le deuxième domaine catalytique, à savoir la triade Asp/Glu/Asp située aux positions respectives 2210/2248/2322 de la séquence ID No. 2, représentée dans la  
25 figure 8.

Ainsi, l'invention porte également sur tout polypeptide isolé ayant une activité catalytique de glycosyltransférase apte à former des dextranes ou des oligosaccharides ayant des ramifications  $\alpha(1\rightarrow2)$  tel qu'obtenu par modification, substitution, insertion ou délétion de séquences  
30 d'acides aminés mais comportant des séquences présentant au moins 80 %

Une structure primaire particulière de l'enzyme selon l'invention est représentée dans la séquence ID No. 2 qui représente une séquence de 2835 acides aminés d'une dextrane-saccharase de *L. mesenteroides* B1299.

5 Cette dextrane-saccharase, nommée DSR-D, possède comme la plupart des glycosyltransférases et des dextrane-saccharases une séquence signal, une région variable faiblement conservée, un domaine catalytique hautement conservé (CD1), un domaine de liaison au glucane (GBD) et un deuxième domaine catalytique (CD2) dans la partie carboxylique de la protéine.  
10 DSR-D est la première glycosyltransférase analysée et présentant deux domaines catalytiques, dans la configuration présentée dans la figure 1 b). C'est également la première glycosyltransférase dont un domaine catalytique est situé dans la partie carboxylique de la protéine.

15 La figure 1b fait apparaître également que le domaine de liaison au glucane est sensiblement plus long que celui décrit précédemment pour les dextranes saccharases connues ; ainsi, une autre caractéristique des enzymes selon l'invention est la taille de ce domaine qui est supérieur à 500 amino-acides.

20 La comparaison et l'analyse de la séquence de DSR-D avec les séquences des glycosyltransférases ou des dextrane-saccharases déjà décrites (1), ainsi que les moyens utilisés à cette fin sont indiqués dans l'exemple 2 détaillé ci-après. Il y apparaît clairement que si l'existence de deux domaines catalytiques différencie substantiellement DSR-D des autres enzymes, en revanche les séquences desdits domaines sont substantiellement conservées. En particulier, les acides aminés nécessaires à l'activité catalytique sont conservés dans le  
25 deuxième domaine catalytique, à savoir la triade Asp/Glu/Asp située aux positions respectives 2210/2248/2322 de la séquence ID No. 2, représentée dans la figure 8.

30 Ainsi, l'invention porte également sur tout polypeptide isolé ayant une activité catalytique de glycosyltransférase apte à former des dextranes ou des oligosaccharides ayant des ramifications  $\alpha(1\rightarrow2)$  tel qu'obtenu par modification, substitution, insertion ou délétion de séquences d'acides aminés mais comportant des séquences présentant au moins 80 %

et de préférence au moins 90 % de similarité avec les séquences suivantes de la séquence ID No. 2 :

423 – 439	2120 – 2138
478 – 501	2161 – 2184
519 – 539	2202 – 2214
560 – 571	2243 – 2250
631 – 645	2315 – 2322
1014 – 1021	2689 – 2696

De façon préférée, enfin, un polypeptide à activité catalytique selon l'invention contient les acides aminés suivants :

W en positions 425 et 2122,

E en positions 430, 565 et 2127, 2248

D en positions 487, 489, 527, 638, 2170, 2172, 2210 et 2322,

H en position 637 et 2321

Q en position 1019 et 2694.

Les polypeptides à activité de glycosyltransférases aptes à former des liaisons osidiques  $\alpha(1\rightarrow2)$  peuvent se présenter sous forme isolée, ou au contraire intégrés dans une protéine plus large, comme par exemple une protéine de fusion. Il peut être en effet avantageux d'inclure des séquences présentant une autre fonction, comme par exemple une séquence étiquette spécifique d'un ligand permettant d'en faciliter la purification. Ces séquences étiquettes peuvent être du type GST (glutathion-S-Transférase), Intéine - CBD (Chitine-Binding Domaine), (commercialisé par New England Biolabs, <http://www.neb.com>), MBD (Maltose Binding Domain), polypeptides contenant des résidus histidine contigus permettant de faciliter la purification du polypeptide avec lequel il est fusionné. L'homme du métier peut concevoir toute autre protéine de fusion permettant d'associer la fonction de la DSR-D de l'invention avec une autre fonction, comme par exemple, et sans être limitatif, une

séquence augmentant la stabilité de l'enzyme produite par expression dans un hôte recombinant ou une séquence apte à augmenter la spécificité ou l'efficacité d'action de cette enzyme, ou une séquence visant à associer une autre activité enzymatique connexe.

De telles protéines de fusion font également partie de l'invention dès lors qu'elles contiennent le domaine CD<sub>2</sub> et le site de liaison au glucane. De la même façon, les fragments de la séquence ID No. 2, comprenant au moins la séquence ID No. 1 et le domaine de liaison au glucane, seuls ou intégrés dans une séquence polypeptidique plus large font partie de l'invention, à partir du moment où l'activité enzymatique de dextrane-saccharase est conservée.

Les variants des séquences polypeptidiques définies ci-dessus font également partie de l'invention. Outre les polypeptides obtenus par substitution conservative des acides aminés telle que définie plus haut, les variants incluent des polypeptides dont l'activité enzymatique est améliorée par exemple par mutagenèse dirigée ou aléatoire, par évolution moléculaire, ou par duplication du domaine catalytique CD<sub>2</sub>.

La structure particulière de cette enzyme identifiée dans la présente invention résulte d'un processus comprenant :

- a) l'identification et l'isolement de la dextrane-saccharase de *L. mesenteroides* catalysant la production des GOS d'intérêt portant les ramifications  $\alpha(1\rightarrow2)$  ;
- b) le séquençage de fragments de l'enzyme ;
- c) la synthèse d'amorces d'amplification aptes à amplifier le gène correspondant de la souche productrice ou des fragments de ceux-ci ;
- d) le séquençage des fragments amplifiés ;
- e) le clonage dans des vecteurs spécifiques et leur expression dans des hôtes appropriés.

Les modalités du procédé mis en œuvre sont détaillées dans la partie expérimentale ci-après. La première étape consiste en une séparation des protéines par électrophorèse en gel de polyacrylamide, et

identification des bandes présentant l'activité de dextrane saccharase par une réaction enzymatique *in situ* en présence de substrat et d'accepteur. La nature des GOS synthétisés est ensuite identifiée sur chaque bande par analyse HPLC selon les méthodes décrites dans (1). Le temps de rétention des oligosides en HPLC dépend de la nature et de l'organisation de leurs liaisons osidiques. Il est possible en particulier de distinguer ceux constitués de résidus liés en  $\alpha(1\rightarrow6)$ , en  $\alpha(1\rightarrow6)$  avec une ramification  $\alpha(1\rightarrow2)$  à l'extrémité non réductrice de la molécule, et ceux recherchés composés d'une chaîne linéaire  $\alpha(1\rightarrow6)$  avec des ramifications  $\alpha(1\rightarrow2)$ .

Les inventeurs ont donc isolé et identifié la dextrane-saccharase de *L. mesentéroïdes* NRRL B-1299 productrice des GOS d'intérêt.

Un procédé d'ingénierie reverse mis en oeuvre dans les étapes b) à e) ci-dessus a permis ensuite de fournir la séquence nucléotidique codant l'enzyme et permettant de la produire en quantité industrielle et le cas échéant de la modifier, d'en améliorer ses performances par les techniques à la disposition de l'homme du métier. A titre d'exemple, on peut citer la mutagenèse dirigée ou aléatoire, ou l'évolution moléculaire (DNA shuffling) (3).

Un autre aspect de l'invention porte sur une molécule d'acide nucléique isolée codant une enzyme à activité glycosyltransférase apte à former des dextrans ou des oligosides présentant des ramifications  $\alpha(1\rightarrow2)$  et comprenant au moins une séquence codant un domaine de liaison au glucane, et au moins une séquence nucléotidique codant un domaine catalytique situé en 3' de la précédente, ladite séquence codant un domaine catalytique ayant au moins 50 % et de préférence au moins 70 % de similarité avec la séquence ID No. 3 représentée dans la figure 9.

Par similarité, on entend le fait que, pour un même cadre de lecture, un triplet donné est traduit par le même acide aminé. Ce terme inclut donc les modifications de bases résultant de la dégénérescence du code génétique.

Le pourcentage de similarité est déterminé en comparant une séquence donnée avec la séquence de référence. Lorsque celles-ci sont de longueurs différentes, le pourcentage de similarité est basé sur le pourcentage de nucléotides de la séquence la plus courte similaires à ceux de la séquence la plus longue.

Le degré de similarité peut être déterminé conventionnellement par utilisation de logiciels tels le ClustalW (Thompson et al., Nucleic Acids Research 22 (1994), 4673-4680) distribués par Julie Thompson ([Thompson@EMBL-Heidelberg.DE](mailto:Thompson@EMBL-Heidelberg.DE)) et Toby Gibson ([Gibson@EMBL-Heidelberg.DE](mailto:Gibson@EMBL-Heidelberg.DE)) du Laboratoire Européen de Biologie Moléculaire, Meyerhofstrasse 1, D 69117 Heidelberg, Germany. ClustalW peut aussi être chargé à partir de plusieurs sites web incluant IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire, B.P. 163, 67404 Illkirch cedex France; <ftp://ftp-igbmc.u-strabg.fr/pub/>) et EBI (<ftp://ftp.ebi.ac.uk/pub/software/>) et tous les sites renvoyant à l'Institut de Bioinformatique, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK).

Les acides nucléiques isolés selon l'invention peuvent comprendre notamment d'autres séquences destinées à améliorer l'expression et/ou l'activité de l'enzyme produite.

Il peut s'agir à titre d'exemple :

- des séquences codant une séquence signal pour leur sécrétion ;
- une duplication de la séquence codant le domaine catalytique CD<sub>2</sub>.

De façon préférée, un acide nucléique isolé selon l'invention comprend :

- a) deux séquences codant des domaines catalytiques ayant au moins 50 %, et de préférence au moins 80 % de similarité avec la séquence ID n° 3 ;



b) une séquence codant le domaine de liaison au glucane, cette dernière étant située de préférence entre les deux séquences en a).

Un acide nucléique selon l'invention pourra comprendre en outre :

5 - un promoteur, apte à son expression dans une cellule hôte choisie,

- une séquence codant un peptide signal, et/ou

- une ou des séquences variables,

cette ou ces séquence(s) étant toutes situées en partie 5' des séquences codant le ou les domaine(s) catalytique(s).

10 Un exemple particulier d'un acide nucléique isolé selon l'invention comprend plus particulièrement :

a) la séquence ID No. 4 représentée dans la figure 10,

b) une séquence présentant au moins 80 % de similarité avec la séquence ID n° 4, ou

15 c) le brin complémentaire de la séquence a) ou b), ou

d) une séquence hybridant a), b) ou c).

L'hybridation en d) est réalisée en conditions standard, et de préférence en conditions stringentes. Par hybridation en condition stringente, on entend le fait qu'il existe une identité de séquences d'au moins 80 % de la séquence que l'on cherche à hybrider et de préférence une identité d'au moins 90 % de la séquence que l'on cherche à hybrider, dans des conditions décrites par exemple dans Sambrook et al. (3<sup>ème</sup> édition, 2001, Coll. Spring Harbour, Laboratory Press, Coll. Spring Harbour, NY).

20 L'invention porte également sur un gène codant une dextrane-saccharase apte à former au moins 15 % de ramifications  $\alpha(1\rightarrow2)$ . Outre la séquence codante, le gène comprend les séquences permettant l'initiation de la transcription ainsi que les séquences permettant l'attachement de l'ARN messager au ribosome (RBS). La séquence ID No. 5 représentée

dans la figure 11 représente une structure du gène tel qu'isolé de *L. mesenteroides* NRRL B-1299.

Les nucléotides en amont de l'ATG d'initiation de la traduction sont numérotés 1 à 232.

5 On peut identifier l'existence d'une séquence RBS entre les nucléotides 218 et 223, ainsi que les séquences consensus - 35 et - 10 situées entre les nucléotides 82 et 86 (TTGAA), d'une part et 100 et 105 (ATAAT), d'autre part.

10 Toute séquence d'acide nucléique hybridable avec l'ADN de la séquence ID No. 4 ou son brin complémentaire est susceptible de coder une enzyme ayant les propriétés et caractéristiques de l'enzyme selon l'invention. Ceci s'applique tant aux séquences naturelles existant dans d'autres micro-organismes que *L. mesenteroides* NRRL-1299 et isolées de banques génomiques de micro-organismes, que celles préparées par génie  
15 génétique ou par synthèse chimique.

En particulier, les séquences en amont de l'ATG d'initiation de la traduction et nécessaires à l'expression de la protéine peuvent être avantageusement substituées par des séquences d'initiation de la transcription et/ou de fixation au ribosome adaptés au système  
20 d'expression choisi pour la séquence codante.

Une séquence d'acides nucléiques susceptible de s'hybrider en condition stringente avec l'acide nucléique isolé selon l'invention comprend également des fragments, des dérivés, ou des variants alléliques de la séquence d'acides nucléiques selon l'invention qui code une protéine  
25 ayant l'activité enzymatique décrite ci-avant. Ainsi, les fragments sont définis comme des fragments de molécules d'acides nucléiques suffisamment longs pour coder une protéine ayant conservé son activité enzymatique. Celle-ci inclut aussi bien des fragments dépourvus de la séquence codant le peptide signal responsable de la sécrétion de la  
30 protéine.

Le terme "dérivé" signifie séquence, différente de la séquence originelle, à une ou plusieurs positions, mais présentant un haut degré de similarité avec ces séquences. Dans ce contexte, similarité signifie une identité d'au moins 80 % des nucléotides, et de préférence d'au moins 90 % avec la séquence originelle. Les modifications dans ce cas portent sur des délétions, substitutions, insertions ou recombinaisons, à partir du moment où l'enzyme codée par ces séquences homologues présentent l'activité enzymatique des polypeptides selon l'invention.

Les séquences d'acides nucléiques selon l'invention telles que décrites ci-dessus et qualifiées de dérivés de ces molécules telles que définies ci-avant, sont généralement des variants exerçant la même fonction biologique. Ces variations peuvent être des variations naturelles, notamment celles observables d'une espèce à l'autre et résultant d'une variabilité inter espèce ou au contraire être introduites par le moyen d'une mutagenèse dirigée, aléatoire ou par DNA Shuffling (évolution moléculaire).

De la même façon, font partie de l'invention les acides nucléiques isolés codant une glycosyltransférase apte à catalyser la synthèse de dextrane ou d'oligosaccharide portant au moins 20 % et de préférence au moins 30 % de ramifications de type  $\alpha(1 \rightarrow 2)$  et obtenus par évolution moléculaire (DNA shuffling) et comprenant :

- une étape de modification aléatoire d'une des séquences décrites précédemment et, en particulier, des séquences ID No. 3 et 4 et d'établissement de variants ;

- une étape d'expression de ces séquences modifiées dans une cellule hôte appropriée, un hôte abritant un variant ;

- une étape de criblage des hôtes exprimant une enzyme apte à former plus de 20 % et de préférence plus de 30 % de liaisons  $\alpha(1 \rightarrow 2)$  sur un substrat approprié et une étape d'isolement du ou des gènes améliorés.

Un acide nucléique isolé selon l'invention pourra également comprendre :

a) une séquence ayant au moins 80 % de similarité avec la séquence codant une dextrane-saccharase exprimée par le plasmide pCR-T7-dsr D dans *E. coli* déposé à la CNCM le 15 mars 2001 sous le numéro I-2649 (*E. coli* TM 109 [pCR-T7-dsrD]), ou

5 b) une séquence complémentaire de la séquence en a).

L'invention porte également sur les fragments d'acides nucléiques tels que définis ci-dessus, hybridables avec la séquence ID No. 4, et utilisables comme sondes d'hybridation pour la détection de séquences codant des enzymes selon l'invention. Ces fragments peuvent  
10 être préparés par toutes les techniques connues de l'homme du métier.

Outre les sondes d'hybridation, des amorces d'amplification font également partie de l'invention. Lesdites amorces sont des fragments hybridables avec la SEQ ID No. 4 ou avec son brin complémentaire et permettent l'amplification de séquences spécifiques codant des dextrane-saccharases présentes dans un organisme procaryote ou eucaryote,  
15 animal ou végétal.

L'utilisation de telles amorces d'amplification permet la mise en oeuvre d'un procédé d'identification de l'existence éventuelle d'un gène codant une enzyme apte à catalyser la synthèse de GOS avec des ramifications  $\alpha(1\rightarrow2)$  dans un tel organisme, ledit procédé faisant  
20 également partie de l'invention.

L'invention porte également sur des vecteurs d'expression comprenant un acide nucléique tel que décrit ci-avant, sous le contrôle de séquence permettant son expression et de préférence son excrétion dans  
25 des cellules procaryotes ou eucaryotes. Par cellules procaryotes, on choisira de préférence des bactéries choisies dans un groupe comprenant *E. coli*, les *Lactococcus*, les *Bacillus*, les *Leuconostoc*. Par cellules eucaryotes, on choisira de préférence les eucaryotes choisis dans un groupe contenant les levures, les champignons ou les végétaux.

30 Le vecteur comprend un promoteur adapté à l'expression de l'acide nucléique isolé selon l'invention dans le système d'expression choisi.

A titre d'exemple, le promoteur du bactériophage T7 pourrait être avantageusement choisi pour une expression dans *E. Coli*.

L'invention porte également sur les cellules hôtes, procaryotes ou eucaryotes, transformées par un acide nucléique selon l'invention de préférence compris dans un vecteur d'expression portant un promoteur, adapté à une expression dans les cellules hôtes choisies. Les cellules transformées sont choisies dans le groupe des bactéries à Gram- telle *E. coli*, ou dans le groupe des bactéries à Gram+ telles *Lactococcus*, *Bacillus*, *Leuconostoc* ou parmi les eucaryotes dans un groupe comprenant les levurés ou les champignons, ou les végétaux.

Un exemple particulier d'une cellule transformée selon l'invention est la souche *E. coli* porteuse d'un plasmide appelé PCR-T7dsrD et porteur de la séquence ID No. 4 sous le contrôle du promoteur du bactériophage T7 et déposée à la CNCM le 15 mars 2001 sous le numéro I-2649.

La présente invention, par ailleurs, porte sur un procédé de production d'une glycosyltransférase apte à former des dextrans ou des oligosides présentant au moins 15 % et de préférence au moins 20 % de ramifications de type  $\alpha(1\rightarrow2)$  osidiques et comprenant :

a) l'insertion d'un acide nucléique ou d'un vecteur tel que décrit précédemment dans une cellule hôte apte à l'exprimer et de préférence à sécréter la glycosyltransférase ;

b) la caractérisation de l'activité enzymatique recherchée par toutes les méthodes accessibles à l'homme du métier ;

c) la purification de l'enzyme à partir d'un extrait cellulaire.

Par méthode de caractérisation de l'activité enzymatique connue de l'homme du métier, on comprendra les méthodes décrites dans la littérature par exemple dans la référence (2) ainsi que de nouvelles méthodes susceptibles d'être mises au point permettant d'identifier et de discriminer les glucooligosaccharides présentant le taux de ramification recherché.

Il s'agit en fait de tout procédé de criblage permettant d'identifier la présence de ramifications  $\alpha(1\rightarrow2)$  dans un GOS.

A titre d'exemple, seront utilisées :

- l'HPLC pour lequel la migration des GOS varie en fonction de la nature et le positionnement des ramifications, notamment ceux ayant le lien  $\alpha(1\rightarrow2)$  à l'extrémité réductrice et ceux ayant ce lien sur l'avant-dernier glucose, et/ou

- la Résonnance magnétique nucléaire (RMN),

- l'existence d'une réaction positive avec des anticorps monoclonaux spécifiques des liaisons  $\alpha(1\rightarrow2)$  sur l'extrémité réductrice et/ou d'anticorps monoclonaux spécifiques des liaisons  $\alpha(1\rightarrow2)$  sur l'avant-dernier glucose du GOS.

L'invention porte également sur un procédé d'obtention d'une glycosyltransférase apte à présenter des oligosides ou des dextrans présentant un taux de ramification  $\alpha(1\rightarrow2)$  supérieur à 15 % et de préférence supérieur à 30 % de la totalité des liaisons osidiques et comprenant une étape de modification de la séquence ID No. 4 par addition, délétion, mutation à partir du moment où :

- le cadre de lecture n'est pas modifié, et

- les acides aminés suivants sont conservés après traduction :

W en positions 425 ou 2122, codé par le triplet TGG en positions 1273 et 6364,

E en positions 430, 565, 2127 et 2248 codés par les triplets GAA en positions 1288, 1693, 6379 et 6742 respectivement,

D en positions 487, 489, 527, 638, 2170 et 2210 codés par les triplets GAT en positions 1459, 1465, 1579, 1912, 6508 et 6628 respectivement,

D en positions 2172 et 2322 codés par les triplets GAT en positions 6514 et 6964,

H en position 637 et 2321, codés respectivement par les triplets CAT en position 1909 et CAC en position 6961,

Q en positions 1019 et 2694 codés respectivement par les triplets CAA (3055) et CAG (8080).

Un procédé de production d'une glycosyltransférase selon l'invention ayant les mêmes caractéristiques que ci-avant peut également comprendre :

- une étape de modification aléatoire de la séquence ID n° 4 et d'établissement d'une banque de variants,

- une étape d'expression de ces séquences modifiées dans une cellule hôte appropriée, un hôte abritant un variant,

- une étape de criblage des hôtes exprimant une enzyme apte à former plus de 15 % et de préférence plus de 30 % de liaison  $\alpha(1 \rightarrow 2)$  sur un substrat approprié,

- une étape d'isolement du ou des gènes améliorés.

Dans un autre mode de réalisation de l'invention, le procédé consiste à modifier la séquence ID No. 3 par duplication de tout ou partie du domaine catalytique CD<sub>2</sub>.

On pourra comprendre que les procédés ci-dessus visent non seulement à l'obtention d'une glycosyltransférase apte à former des oligosides présentant un taux de ramification  $\alpha(1 \rightarrow 2)$  constant et reproductible, supérieur à 15 % des ramifications totales mais également à améliorer le taux de ramification  $\alpha(1 \rightarrow 2)$  dans l'objectif de modifier les propriétés des oligosides obtenus dans le sens d'une amélioration de leurs propriétés diététiques ou de leur capacité à maintenir ou rétablir la flore bactérienne associée à certains organes du corps humain ou animal.

La présente invention porte enfin sur les glycosyltransférases susceptibles d'être obtenues par un procédé cité ci-avant et apte à former au moins 15 % et de préférence au moins 30 % de ramifications de type  $\alpha(1 \rightarrow 2)$  osidiques dans des glucooligosaccharides.

L'invention porte enfin sur l'utilisation des glycosyltransférases selon l'invention ainsi que celles susceptibles d'être obtenues par les procédés ci-dessus, dans la fabrication d'une composition à effet

prébiotique ou dans la fabrication d'une composition dermatologique, cosmétique ou pharmaceutique.

A titre d'exemples non limitatifs, on peut citer l'amélioration du transit intestinal chez les animaux et chez l'homme, l'amélioration de l'assimilation du calcium et/ou du magnésium et des minéraux en général, la prévention du cancer du colon, la prévention ou le traitement des affections de la peau telles l'acné, les pellicules, les odeurs corporelles.

L'avantage des polypeptides et des acides nucléiques codant ces polypeptides selon l'invention se situe non seulement au niveau de l'amélioration en terme de qualité, de rendement, de reproductibilité, et de prix de revient des glycosyltransférases aptes à former des oligosaccharides avec des ramifications de type  $\alpha(1 \rightarrow 2)$  osidiques mais également dans la perspective de produire de nouvelles enzymes dont la fonctionnalité est améliorée.

Les figures, exemples et description détaillés ci-après permettent, sans la limiter, d'illustrer les caractéristiques et les fonctionnalités particulières des polypeptides à activité enzymatique et des séquences codant ceux-ci. Elles permettent en particulier d'illustrer de façon plus précise la spécificité du domaine catalytique présent dans la partie carboxylique de l'enzyme selon l'invention et son évolution potentielle pour l'obtention d'enzymes améliorées.

#### **LEGENDE DES FIGURES :**

**Figure 1 :** structure des glycosyltransférases : la figure 1a) représente la structure des glycosyltransférases et des dextrane-saccharases décrites dans la littérature (1). La figure 1b) représente la structure de la glycosyltransférase selon l'invention. A : peptide signal ; B : région variable, C : domaine catalytique, D : domaine de liaison au glucane (GBD).

**Figure 2 :** schéma récapitulatif de la méthode de clonage de la séquence nucléotidique codant une glycosyltransférase selon l'invention à



l'aide d'une bibliothèque génomique en utilisant une sonde PCR décrite dans le tableau I et une sonde HindIII/EcoRV respectivement.

Figure 3 : comparaison des séquences signal de différentes glycosyltransférases de *L. mesenteroides*. Les acides aminés conservés sont en gras. DSR-B : *L. mesenteroides* B-1299 (4) ; DSR-S : *L. mesenteroides* B-512F (5) ; ASR : *L. mesenteroides* B-1355 (6).

Figure 4 : alignement des 11 séquences répétées de l'enzyme DSR-D et observées dans la zone variable.

Figure 5 : alignement des séquences conservées du domaine catalytique.

- Bloc A : acides aminés essentiels de la partie N-terminale du domaine catalytique ;

- Bloc B : acides aminés de la partie du domaine catalytique de liaison au saccharose ;

- Blocs C, D, E : blocs contenant les trois résidus d'acides aminés impliqués dans la triade catalytique (6) ;

- Bloc F : séquence contenant la glutamine 937 de GTF-I étudiée par Monchois et al. (7).

Les acides aminés entièrement conservés sont indiqués en gras. \*\* : substitutions conservatives ; \* : substitutions semi-conservatives ; — : GAP. Les numérotations sont celles de la séquence ID No. 2.

Figure 6 : caractérisation HPLC des produits synthétisés par l'enzyme recombinante DSR-D.

6A : analyse en HPLC des glucooligosaccharides obtenus avec les dextrane-saccharases de *L. mesenteroides* NRLL B-1299.

6B : analyse HPLC des glucooligosaccharides obtenus par la DSR-D recombinante. L'identification des différents pics suivants :

1 : fructose,

2 : maltose,

3 : sucrose,

4 : panose,

5 : R4,

6 : OD4,

7 : R5,

8 : OD5, -

A, B, C : pics non identifiés.

6C : DSR-D recombinante délétée du domaine catalytique de la partie carboxylique de l'enzyme ( $\Delta$  DSR-D).

Figure 7 : séquence peptidique de CD<sub>2</sub>.

Figure 8 : séquence peptidique de DSR-D.

Figure 9 : séquence nucléotidique codant CD<sub>2</sub>.

Figure 10 : séquence nucléotidique codant DSR-D.

Figure 11 : séquence nucléotidique d'un gène codant DSR-D.

#### MATERIELS ET METHODES :

##### 1) Souches bactériennes, plasmides et conditions de croissance :

Toutes les souches sont conservées à  $-80^{\circ}\text{C}$  dans des tubes contenant 15% de glycérol (v/v).

*Leuconostoc mesenteroides* B-1299 (NRRL, Peoria, USA) est cultivée à  $27^{\circ}\text{C}$ , sous agitation (200 RPM) sur milieu standard (saccharose 40 g.l<sup>-1</sup>, phosphate potassium 20 g.l<sup>-1</sup>, extrait de levure 20 g.l<sup>-1</sup>, MgSO<sub>4</sub>·7H<sub>2</sub>O 0.2 g.l<sup>-1</sup>, MnSO<sub>4</sub>·H<sub>2</sub>O 0.01g.l<sup>-1</sup>, NaCl 0.01g.l<sup>-1</sup>, CaCl<sub>2</sub> 0.02 g.l<sup>-1</sup>, FeSO<sub>4</sub>·7H<sub>2</sub>O 0.01 g.l<sup>-1</sup>), le pH étant ajusté à 6,9.

*Escherichia coli* DH5 $\alpha$  et JM109 ont été cultivées sur milieu LB (Luria-Bertani).

La sélection des clones recombinants de pUC18 ou pGEM-T Easy est effectuée sur boîtes LB-agar supplémenté avec 100  $\mu\text{g.ml}^{-1}$  d'ampicilline, 0.5 mM d'isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) et 40  $\mu\text{g.ml}^{-1}$  de 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-gal). Des cellules d' *E. coli* TOP 10 ont été utilisées pour le système de clonage de produit PCR

TOPO Cloning (Invitrogen), et cultivées sur milieu LB supplémenté de kanamycine à la concentration de  $50 \mu\text{g.ml}^{-1}$ .

En ce qui concerne l'expression de *dsrD*, le kit de clonage ECHO Cloning System (Invitrogen) permet le clonage d'un produit PCR dans un vecteur donneur (pUNI/V5-His-TOPO), précédant une étape de recombinaison avec un vecteur accepteur adapté (pCR-T7-E). Ce système requiert des cellules *E. coli* PYR1, TOP 10 et BL21(DE3)pLysS cultivées sur milieu LB supplémenté de  $50 \mu\text{g.ml}^{-1}$  de kanamycine, ainsi que de  $34 \mu\text{g.ml}^{-1}$  de chloramphenicol pour la souche BL21(DE3)pLysS.

Les plasmides pUC18 digérés et déphosphorylés proviennent de Pharmacia (Amersham Pharmacia Biotech) et ont été utilisés pour la constitution de banque d'ADN génomique de *L. mesenteroides* B-1299. Le clonage de produit PCR a, quant à lui, nécessité l'emploi de plasmide pGEM-T Easy (Promega), et pour les fragments de plus de 2 kbp de plasmide TOPO-XL (Invitrogen)

## 2) Electrophorèse sur gel, localisation et caractérisation de l'enzyme :

Après une culture de *L. mesenteroides* B-1299 de 7h, le milieu est centrifugé (7000 RPM,  $4^{\circ}\text{C}$ , 30 min) et les cellules, où 90% de l'activité enzymatique se retrouve, sont concentrées 10 fois dans une solution de tampon acétate (20 mM, pH 5,4), chauffées 5 minutes à  $95^{\circ}\text{C}$  en présence de solution de dénaturation (Tris HCl 62.5 mM, SDS 4%, urée 6M, bleu de bromophenol 0.01% et  $\beta$ -mercaptoethanol 200 mM). 300  $\mu\text{l}$  du mélange est déposé sur gel de polyacrylamide à 7%. Après migration, les protéines totales sont révélées par coloration au noir amido, alors que l'activité dextrane-saccharase est détectée par coloration du polymère au réactif de Schiff après synthèse de dextrane *in situ*. Les bandes correspondant à des dextrane-saccharases actives sont excisées et incubées séparément dans 2 ml de solution d'acétate de sodium 20 mM pH 5.4 contenant  $100\text{g.l}^{-1}$  de saccharose et  $50 \text{g.l}^{-1}$  de maltose. Après consommation totale du saccharose, la réaction est arrêtée par chauffage à

95°C pendant 5 minutes, et le milieu réactionnel est centrifugé 5 minutes à 15000g afin d'éliminer le dextrane insoluble. Les échantillons sont analysés par chromatographie en phase inverse (colonne C18, Ultrasep 100; 6 µm, 5x300mm, Bishoff Chromatography) en utilisant de l'eau ultrapure comme éluant, à un débit constant de 0.5ml.min<sup>-1</sup>. Les oligosaccharides sont séparés pendant 30 minutes à température ambiante, et détectés par réfractométrie. Le séquençage peptidique a été réalisée sur les bandes protéiques sélectionnées par le Laboratoire de Microséquençage, Institut Pasteur, Paris.

### 3) Techniques de biologie moléculaire utilisée

La purification du plasmide d' *E. coli* et la purification du DNA génomique de *L. mesenteroides* ont été réalisées en utilisation respectivement QiaPrep Spin Plasmid kit et le Cell Culture DNA maxi kit (QiaGen). Les procédés d'amplification et de clonage ont été réalisés en utilisant les techniques standards (Sambrook et al. 2001, voir supra). Les enzymes de restriction et de modification proviennent des sociétés commerciales New England Biolabs ou Gibco BRL et utilisées selon les protocoles des fabricants.

La PCR est réalisée avec des amorces choisies sur la base de la séquence protéique obtenue sur une bande de gel d'électrophorèse isolée (voir plus haut, électrophorèse sur gel et localisation de l'enzyme). Deux peptides ont été sélectionnés :

- 29-FYFESGK, et
- 24-FESQNNNP

et utilisés pour synthétiser des oligonucléotides dégénérés et indiqués dans le tableau I ci-dessous.

Dans ce tableau où les numérotations sont celles de la figure 10, il apparaît que la présence d'un résidu sérine dans les deux peptides nécessite la synthèse de deux amorces pour chaque peptide dans la mesure où la sérine peut être codée par six codons différents. ECHO-dir et ECHO-inv sont les amorces utilisées ayant permis l'amplification de *dsrD*

par PCR pour son clonage dans le système d'expression Echo Cloning (Invitrogen).

TABLEAU I :

5

Désignation	Description	Séquence 5'-3'
29-dir1	FYFESGK	TT(C/T)TA(C/T)TT(C/T)GA(A/G)TCAGG(C/G)AA(A/G)
29-dir2		TT(C/T)TA(C/T)TT(C/T)GA(A/G)AGCGG(C/G)AA(A/G)
24-inv1	FESQNNP	(T/G)GG(G/A)TT(G/A)TT(G/A)TTTTGTGA(T/C)TCAAA
24-inv2		(T/G)GG(G/A)TT(G/A)TT(G/A)TTTTGGCT(T/C)TCAAA
IPCR-rev	séquence nt 5769-5798	CCCTTTACAAGCTGATTTTGCTTATCTGCG
IPCR-dir	séquence nt 8311-8342	GGGTCAAATCCTTACTATACATTGTCACACGG
ECHO-dir	séquence nt -6 - 39	AGTTGTATGAGAGACATGAGGGTAATTTGTGACCGTAAAAAATTG
ECHO-inv	séquence nt 8457-8504	ATTTGAGGTAATGTTGATTTATCACCATCAAGCTTGAAATATTGACC

#### PCR :

10 La PCR est réalisée en utilisant un thermocycleur Perkin-Elmer, modèle 2400, et en utilisant 50 nanogrammes de l'ADN génomique. Les quantités d'amorces utilisées sont de 10 µM de 29-Dir1 et de 24-Inv1. Au mélange réactionnel, sont ajoutés 250 µM de chaque désoxynucléotide triphosphate, et la Taq Polymérase.

15 Après une amplification de 25 cycles à 94° C pendant 30 secondes puis à 50° C pendant 30 secondes, puis à 72° C pendant 5 minutes, un fragment de 666 paires de base a été obtenu.

**Hybridation southern et bibliothèque génomique de *L. mesenteroides* B-1299 :**

L'ADN chromosomique de *L. mesenteroides* B-1299 a été digéré avec différentes enzymes de restriction, puis séparé sur gel d'agarose. Des bibliothèques génomiques de la bactérie ont été transférées sur des membranes de nylon hybond N+ (Amersham PharmaciaBiotech). L'hybridation a été réalisée en utilisant le fragment de 666 paires de bases à la désoxy-adénosine-triphosphate marqué au  $^{32}\text{P}$ . La réaction de marquage a été réalisée en utilisant le kit de marquage "Mega Prime DNA Labelling System DIT" (Amersham PharmaciaBiotech), suivie par la purification de la sonde sur des colonnes MicroSpin S-200HR. La pré-hybridation et l'hybridation ont été réalisées en conditions fortement stringentes (65° C pendant la nuit, selon les méthodes habituelles) (Maniatis et al., 2001).

**PCR inverse :**

L'ADN génomique de *L. mesenteroides* B-1299 est digéré par *EcoRV* dans les conditions recommandées par le fournisseur.

Après re-circularisation, les produits de digestion sont utilisés comme matrice dans une PCR inverse (Extrapol II DNA polymérase (Eurobio) 25 cycles, 94° C, 30 secondes ; 51° C, 30 secondes ; 72° C, 3 minutes). Les deux amorces ont été choisies en fonction de la séquence de l'insert de pSB2 comme ceci est indiqué dans la figure 2.

La figure 2 résume les modalités d'obtention des différents plasmides porteurs des fragments de *dsrD* par criblage de la bibliothèque génomique et utilisation des sondes décrites ci-dessus.

**Séquence d'ADN et analyse :**

Après le séquençage des peptides, des amorces dégénérées dessinées en tenant compte de la fréquence d'utilisation des codons dans les gènes de dextrane-saccharases de *L. mesenteroides* B-1299, ont été synthétisées et ont permis l'amplification d'un fragment de 666 bp. Le

séquençage de ce fragment a révélé de fortes homologues avec les gènes de dextrane-saccharases déjà connus, tout en étant totalement nouveau.

L'utilisation de ce fragment comme sonde homologuée dans des expériences de Southern, a permis de repérer des signaux positifs sur différentes pistes d'ADN génomique digéré. Une première banque *HindIII* a ainsi été criblée, et un plasmide recombinant, nommé pSB2, contenant un insert de 5,6 kbp, a été purifié. L'analyse de la séquence de ce fragment *HindIII* a révélé un cadre ouvert de lecture couvrant la totalité de l'insert. Ensuite, une banque *EcoRV* a été criblée avec une sonde *HindIII/EcoRV* isolée à l'extrémité N-term de l'insert *HindIII* de 5,6 kbp. Un plasmide recombinant pSB3, testé positivement par dot-blot, s'est avéré contenir un insert de 3,8 kbp qui, après séquençage, a été montré contenir le codon d'initiation de la traduction et la région promotrice du nouveau gène de dextrane-saccharase nommé *dsrD*.

Dans le but d'obtenir le codon de terminaison de *dsrD*, une PCR inverse a été réalisée sur de l'ADN génomique de *L. mesenteroides* B-1299 digéré par *EcoRV* et reliqué sur lui-même, en utilisant des amorces oligonucléotidiques divergentes dessinées à partir de la séquence de l'insert pSB2. Un fragment unique à la taille attendue de 1 kbp a été amplifié puis cloné dans un pGEM-T Easy, pour obtenir le plasmide pSB4. Après séquençage, la séquence amplifiée située en aval du site *HindIII* comporte 221 bp et contient le codon de terminaison du cadre de lecture de *dsrD*, situé 30 bp en aval du site de restriction *HindIII*.

Le séquençage des différents fragments portés par les trois plasmides a été réalisé par la société Génome Express, et ce sur les deux brins. Les analyses des séquences de nucléotides ont été réalisées en utilisant le "ORF Finder" (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), Blast (<http://www.ncbi.nlm.nih.gov/blast/blast.cgi>, Altschul et al., 1997), ClustalW (<http://www2.ebi.ac.uk/clustalw>, Thompson et al., 1994), PRODOM (<http://protein.toulouse.inra.fr/prodom.html>, Corpet et al., 2000), PFAM (<http://pfam.wustl.edu/hmmsearch.shtml>, Bateman et al., 2000) et SAPS

(<http://bioweb.pasteur.fr/seqanal/interfaces/saps.html>, Brendel et al, 1992), l'ensemble de ces logiciels étant accessible par internet.

#### Expression de la protéine :

Le système de clonage ECHO (Invitrogen) a été utilisé : deux amorces telles que proposées dans le tableau I ci-dessus permettent l'amplification en utilisant le système de Boehringer Mannheim "Expand Long Template" dans les conditions suivantes : 94° C pendant 3 minutes suivis de 25 cycles à 94° C pendant 30 secondes, 55° C pendant 30 secondes, et 68° C pendant 7 minutes. Les produits PCR sont ensuite clonés dans le vecteur pUNI/V5-His-TOPO, permettant l'obtention d'un vecteur donneur (pUNI-*dsrD*) pour recombiner avec un vecteur accepteur (pCR-T7-E) et adapté à l'expression dans *E. coli*. Le plasmide final est désigné pCR-T7-*dsrD*.

Cette construction place le gène *dsrD* sous le contrôle du promoteur du bactériophage T7 et permet l'expression inductible du gène *dsrD*.

Après induction avec 1 mM d'IPTG, les cellules d'*E. coli* BL21 transformées sont récoltées par centrifugation après 4 heures de croissance, et re-suspendues à une densité optique finale de 80 à 600 nm dans du tampon acétate de sodium 20 mM pH 5,4 et du Triton X100 à 1% (v/v) en présence de 1 mM PMSF afin d'empêcher la protéolyse dans les extraits cellulaires après sonication.

#### Tests enzymatiques :

Les réactions enzymatiques sont réalisées dans les conditions standards à 30 degrés dans du tampon acétate de sodium 20 mM pH 5,4,  $\text{CaCl}_2$  0,05 g/l<sup>-1</sup>,  $\text{NaN}_3$  1 g/l<sup>-1</sup> et saccharose 100 g/l<sup>-1</sup>. L'activité de l'enzyme DSR-D est déterminée en mesurant la libération des sucres réducteurs, une unité étant définie comme la quantité d'enzyme qui catalyse la formation d'1  $\mu\text{mol}$  de fructose par minute dans les conditions standards. Les oligosaccharides sont synthétisés dans un milieu réactionnel contenant 100 g/l de maltose, 200 g/l de saccharose et 0,5 unités/ml de DSR-D.



Comme pour la synthèse de dextrane, la réaction enzymatique a été poursuivie pendant 24 heures en présence de 100 g/l de glucose. Le dextrane produit a été précipité en présence d'éthanol 50 % (v/v) et lavé deux fois dans l'éthanol à 50 % (v/v) avant lyophilisation. Il est ensuite dissous à 10 mg/ml dans du D<sub>2</sub>O et analysé par spectrométrie deRMrl du <sup>13</sup>C.

#### Analyse par RMN :

Les analyses par résonance magnétique nucléaire (RMN) du proton ont été réalisées sur un spectromètre Varian unit Plus équipé d'un système Ultrashim (fréquence d'opération 499, 836 Mhz). Les échantillons ont été dilués dans de l'eau deutérée (5mg dans 0.75 ml de solvant) et analysés à 25°C. Les spectres de protons ont ensuite été réalisés à une fréquence de 3300Hz pendant 90 pulses et 12480 points. Le temps d'acquisition est égal à 1,891 secondes et le nombre d'accumulations et de 32 scans.

Les spectres du carbone <sup>13</sup>C sont enregistrés à l'aide d'un spectromètre Brüker AC 300. Les échantillons sont dilués dans de l'eau deutérée (environ 10-15 mg dans 0,40 ml). les spectres sont réalisés à une fréquence de 75 MHz à 70°C, et le temps d'acquisition est égal à 0,54 secondes.

Les déplacements chimiques sont donnés en ppm par rapport au signal méthyle de l'acétone dans l'eau, qui est pris comme standard à  $\lambda=31.5$  ppm par rapport au signal méthyle du 4,4 diméthyl-4-silapentane-1-sulfonate.

#### ***Exemple 1 : Caractérisation et purification de l'enzyme***

##### ***DSR-D et obtention du dsrD***

Les enzymes produites par les cultures de *L. mesentéroïdes* et obtenues sur gel de polyacrylamide en SDS tel que décrit dans la partie Matériels et Méthodes sont isolées par découpe du gel.

Les GOS produits par les enzymes ainsi isolées sont analysés par HPLC selon les méthodes décrites dans (1). L'enzyme dont l'activité est recherchée est déduite de la nature des GOS produits. Après protéolyse

trypsique et séparation par HPLC des peptides produits, 2 peptides : 29-FYFESGK et 24- FESQNNNP sont séquencés et utilisés comme modèle pour la synthèse d'amorces nucléotidiques dégénérées.

Les différentes étapes d'amplification et de clonage sont représentées dans la figure 2: Le gène complet est inséré dans le plasmide pCR-T7-E et exprimé dans *E. coli*.

La production d'une enzyme fonctionnelle est attestée par la production des GOS dont l'analyse HPLC est représentée dans la figure 6b).

On remarquera en particulier l'importance des pics 5 et 7 représentatifs des GOS à ramification  $\alpha(1 \rightarrow 2)$ .

### **Exemple 2 : caractérisation des séquences *dsrD* et de *DSR-D***

#### **2.1 Séquence nucléotidique :**

La séquence nucléotidique de l'enzyme est représentée dans la séquence ID No. 4. Elle est composée d'un cadre de lecture de 8508 nucléotides.

La séquence nucléotidique de l'insert dans le plasmide pCR-T7-dsrD contient un site de liaison au ribosome (RBS), 9 bases en amont du codon d'initiation ATG et est composée d'un hexa-nucléotide GAGGAA.

#### **2.2 Analyse de la séquence d'acides-amino :**

La séquence de 8508 nucléotides de *dsrD* code une protéine de 2835 acides aminés et est représentée dans la séquence ID No. 2. Le point isoélectrique de cette protéine est de 4,88 et son poids moléculaire théorique de 313,2 kDa. En dépit des fortes similarités avec les dextrane-saccharases déjà connues, *dsrD* est caractérisée par une structure originale.

L'alignement de la séquence d'acides-amino avec l'ensemble des glycosyltransférases et des dextrane-saccharases connues confirme que la structure en domaine des glycosyltransférases et des dextrane-saccharases est conservée, à savoir : une séquence signal, une région

variable, un domaine catalytique hautement conservé et un domaine de liaison au glucane. Cette structure est représentée dans la figure 1a.

Comme l'indique la figure 1b, un deuxième domaine catalytique forme la partie carboxy-terminale de l'enzyme comme cela été confirmé par PRODOM et une analyse Blast.

Avec un poids moléculaire de 313,2 kDa, DSR-D a environ 2 fois le poids moléculaire moyen des autres glycosyltransférases et dextrane-saccharases (1), ce qui est en accord avec la présence d'un deuxième domaine catalytique à l'extrémité C-terminale et également avec un domaine de liaison au glucane plus long.

a) analyse de la séquence signal :

La séquence signal et la séquence nucléotidique codant le peptide signal sont extrêmement conservées si on les compare aux autres dextrane-saccharases comme ceci est indiqué dans la figure 3. Le site de clivage est localisé entre les acides aminés 40 et 41.

b) domaine variable :

En aval du peptide signal, DSR-D a un domaine variable de 207 acides aminés. Lorsqu'on le compare aux autres domaines variables des glycosyltransférases, en utilisant un programme d'alignement de type SAPS, on met en évidence la présence d'un motif de 14 acides aminés répété 11 fois comme ceci est indiqué dans la figure 4.

Ce motif répété, riche en alanine, threonine et acide aspartique n'a jamais été identifié précédemment.

Le rôle et la signification de cette région n'ont jamais été élucidés. Différentes études ont démontré que sa délétion n'affecte pas l'activité enzymatique (4). Le rôle du motif répété de 14 acides aminés qui n'existe pas dans les autres glycosyltransférases reste néanmoins à déterminer.

c) analyse des domaines catalytiques :

Le premier domaine catalytique s'étend des acides aminés 248 à 1142 (CD1) de la séquence ID No 2, alors que le deuxième est

localisé entre les acides aminés 1980 et 2836 (CD2). Ces deux domaines présentent 45 % d'identité et 65 % de similarité entre eux.

CD1 et CD2 contiennent les acides aminés déjà identifiés dans les glycosyltransférases et les dextrane-saccharases comme étant essentiels à leur activité enzymatique, et comme ceci est indiqué dans la figure 5.

Les triades catalytiques de CD1 et CD2 déterminées par analogie avec l' $\alpha$  amylase (réf. 7) sont présentes aux positions suivantes : (Asp 527/Glu 565/Asp 638 pour CD1 et Asp 2210/Glu 2248/Asp 2322 pour CD2).

D'autres résidus conservés ont été identifiés comme importants pour l'activité enzymatique : les résidus Trp 425/Glu 430 pour CD1 et Trp 2122/Glu 2127 pour CD2, lesquels sont analogues à celles du domaine N-terminal de GFTI décrits par Monchois et al. (4) : Trp 344/Glu 349.

En revanche, certaines séquences situées dans la région conservée des glycosyltransférases et des dextrane-saccharases ne se retrouvent pas dans CD2 de DSR-D. Ainsi, comme indiqué dans le tableau 4 ci-dessus, les séquences FIHNDTI (2214-2220) et KGVQEKV (2323-2329) divergent des autres séquences consensus des dextrane-saccharases déjà étudiées qui sont respectivement NVDADLL et SEVQTVI.

#### d) domaine de liaison au glucane :

Lorsque l'on compare la séquence de DSR-D avec les séquences connues, il apparaît que le domaine de liaison au glucane est sensiblement plus long. En effet, ce domaine a une longueur d'environ 500 acides aminés dans les glycosyltransférases et les dextrane-saccharases étudiées alors que dans DSR-D, il représente 836 acides aminés. Plusieurs motifs répétés A et C ont pu être identifiés. En particulier, une série de répétition AC a pu être identifiée.

### Exemple 3 : expression *dsr-D* dans *E. coli*

Des cellules d'*E. coli* BL21 (DE3) pLysS pCR-T7-*dsrD* ont été cultivées comme décrit ci-dessus. Après gel d'électrophorèse en polyacrylamide-(page-SDS) l'analyse des extraits protéiques révèle effectivement la présence de plusieurs bandes ayant l'activité de dextrane saccharase, ladite activité étant mesurée comme ci-dessus. C'est cette lignée qui a été déposée à la CNCM le 15 mars 2001 sous le numéro I-2649.

#### Identification et caractérisation de l'activité enzymatique :

En utilisant une molécule accepteur de glucose, les dextrane-saccharases produites par *E. coli* recombinant ont été comparées avec celles produites par *L. mesenteroides* B-1299.

L'analyse HPLC des produits de la réaction avec la DSR-D recombinante montre (figure 6) des temps de rétention correspondant aux GOS préalablement identifiés R4 et R5 (2). Les oligosaccharides de type R sont les séries de GOS linéaires, le lien  $\alpha(1 \rightarrow 2)$  étant lié à l'extrémité non-réductrice. La série OD, GOS linéaires résultant de liens glycosidiques  $\alpha(1 \rightarrow 6)$  avec un résidu maltose à l'extrémité réductrice, a été observée en très faibles quantités. Trois nouveaux composés sont en revanche détectés dans les produits de l'enzyme recombinante.

#### Identification des GOS produits :

Finalement, la figure 6b montre clairement que les pics 5 et 7 représentant les GOS de la série R sont relativement plus importants avec l'enzyme recombinante qu'avec l'enzyme native dont les pics correspondant au panose et à OD5 sont plus importants.

### Exemple 4 : Effet de la délétion de *CD2* sur l'activité enzymatique de *DSR-D*

L'ADN génomique de *L. mesenteroides* B-1299 est utilisé comme matrice pour amplifier par PCR le gène *dsrD* délété de la séquence correspondant au second domaine catalytique. Pour cela, 2 oligonucléotides, ECHO-dir (5'-AGTTGTATGAGAGACATGAGGGTAATTTGTGACCGTAAAAAATTG) (SEQ. ID

No. 6), correspondant à la séquence nucléotidique -6 à 39 et contenant le codon d'initiation de la traduction, et ECHO-inv-dél (5'-GTATTAGTGAATAAGTATTCACCATTCGATTATCGTCAAAATAGTACG) (SEQ. ID No. 7) complémentaire de la séquence 5889-5937 et correspondant à la séquence peptidique YYFDDKGNGEYCFTNT, ont été synthétisées, afin de fusionner l'extrémité C-terminale de la protéine déletée avec un tag His présent sur le vecteur de clonage. La réaction PCR a été réalisée grâce à un DNA thermal cycler model 2400 (Perkin-Elmer), avec le système Expand Long Template System (Boehringer Mannheim), suivant le cycle de température: 94°C pendant 3 min, puis 25 cycles avec : 30 s à 94°C, 30s à 55°C et 7 min à 68°C. le produit PCR a ensuite été cloné dans le vecteur donneur pUNI, et le plasmide résultant, utilisé dans une réaction de recombinaison avec le vecteur d'expression pCR-T7- $\Delta$ dsrD.

La préparation des extraits cellulaires, les réactions enzymatiques, l'analyse des produits de la réaction sont les mêmes que dans l'exemple 3 ci-dessus.

Le profil HPLC des GOS obtenus avec l'enzyme DSR-D déletée du domaine CD2 apparaît dans la figure 6 c).

Les GOS de type R, représentés par les pics 5 et 7 visibles dans la figure 6 a) et 6 b) sont totalement absents des produits obtenus avec l'enzyme recombinante déletée de CD2. Les seuls produits analysables sont ceux correspondant à des oligosides linéaires résultant de liens  $\alpha(1 \rightarrow 6)$  avec un résidu maltose dans la partie réductrice. Ce résultat indique clairement le rôle essentiel du domaine catalytique situé dans la partie carboxy-terminale de l'enzyme dans sa capacité à former des liaisons osidiques  $\alpha(1 \rightarrow 2)$ .

En conclusion, les inventeurs ont réussi en isolant une dextrane saccharase particulière produite par *L. mesenteroides* à caractériser une structure particulière et inattendue de cette enzyme apte à produire des oligosides d'intérêt et présentant des ramifications de type

$\alpha(1\rightarrow2)$ . L'identification et la caractérisation de cette séquence permettent d'une part de construire des cellules ou organismes recombinants exprimant de façon spécifique cette enzyme et, d'autre part, d'en envisager sa modification par mutagenèse dirigée ou aléatoire ou par évolution moléculaire (DNA Shuffling) afin d'en améliorer encore ses caractéristiques.

Cette invention permet en outre d'améliorer le rendement et la reproductibilité de la production des GOS d'intérêt pour les différentes applications citées ci-avant.

## REFERENCES BIBLIOGRAPHIQUES

- (1) Monchois V., Willemot R.M., Monsan P. (1999).  
 Glucansucrases : mechanism of action and structure-function relation-  
 ships. FEMS microbiol. Rev. 23, 131-151.
- (2) Dols M., Remaud-Simeon M., Willemot R.M., Vignon  
 M.R., Monsan P.F. (1998). Structural characterization of the maltose  
 acceptor-products synthesised by *Leuconostoc mesenteroides* NRRL B-  
 1299 dextranucrase. Carbohydrate Research 305, 549-559.
- (3) Arnold F.H. (2001). Nature 409 n° 6817, 253.
- (4) Monchois V. Vignon M., Russel R.R.B. (1999). Isolation of  
 key amino-acid residues at the N-terminal end of the core region of  
*Streptococcus downei* glucansucrase GTF-I. Appl. Microbiol. Biotechnol.  
 52, 660-665.
- (5) Wilke-Douglas M., Perchorowicz J.T., Houck C.M.,  
 Thomas B.R. (1989). Methods and compositions for altering physical  
 characteristics of fruit and fruit products. PCT patent, WO 89/12386.
- (6) Arguello-MOrales M.A., Remaud-Simeon M., Pizzut S.,  
 Sarçabal P., Willemot R.M., Monsan P. (2000). Sequence analysis of the  
 gene encoding alternansucrase, a sucrose glucosyltransferase from  
*Leuconostoc mesenteroides* NRRL B-1355. FEMS Microb. Lett. 182, 81-85.
- (7) Devulapalle K.S., Goodman S., Gao Q., Hemsley A.,  
 Mooser G. (1997). Knowledge-based model of a glucosyltransferase from  
 oral bacterial group of mutants *streptococci*. PProtein Sci. 6, 2489-2493.



## REVENDICATIONS

5 1. Polypeptide isolé ayant une activité enzymatique de glycosyltransférase apte à former des dextranes présentant des ramifications  $\alpha(1 \rightarrow 2)$  à partir de saccharose, d' $\alpha$ -D-fluoro-glucose, de para-nitrophényl- $\alpha$ -D-glucopyranoside, d' $\alpha$ -D-glucopyranoside- $\alpha$ -D-sorbofurano-side ou de 4-O- $\alpha$ -D-galactopyranosylsucrose, caractérisé en ce qu'il comprend au moins un domaine de liaison au glucane et un domaine à  
10 activité catalytique situé en aval du domaine de liaison au glucane.

15 2. Polypeptide selon la revendication 1 comprenant au moins deux domaines catalytiques situés de part et d'autre du domaine de liaison au glucane.

3. Polypeptide selon la revendication 1 ou 2 présentant une structure telle que représentée dans la figure 1b.

20 4. Polypeptide selon l'une des revendications 1 à 3 dans lequel le ou les domaine(s) à activité catalytique a (ont) un pourcentage de similarité compris entre 65 % et 100 % avec la SEQ. ID : n°1.

25 5. Polypeptide selon l'une des revendications précédentes dans lequel la taille du domaine de liaison au glucane est supérieure à 500 aminoacides.

6. Polypeptide selon la revendication 5. ayant la SEQ.ID : n°2.

30 7. Polypeptide selon la revendication 6 modifié par substitution, insertion, délétion de séquences d'acides aminés et comportant des séquences présentant au moins 80 % et de préférence au

# REVENDECATIONS

1. Polypeptide isolé ayant une activité enzymatique de glycosyltransférase apte à former des dextrans présentant des ramifications  $\alpha(1 \rightarrow 2)$  à partir de saccharosé, d' $\alpha$ -D-fluoro-glucose, de para-nitrophényl- $\alpha$ -D-glucopyranoside, d' $\alpha$ -D-glucopyranoside- $\alpha$ -D-sorbofurano-side ou de 4-O- $\alpha$ -D-galactopyranosylsucrose, caractérisé en ce qu'il comprend au moins un domaine de liaison au glucane et un domaine à activité catalytique situé en aval du domaine de liaison au glucane.

2. Polypeptide selon la revendication 1 comprenant au moins deux domaines catalytiques situés de part et d'autre du domaine de liaison au glucane.

3. Polypeptide selon la revendication 1 ou 2 comprenant un peptide signal, une région variable, deux domaines catalytiques et un domaine de liaison au glucane situé entre les deux domaines catalytiques.

4. Polypeptide selon l'une des revendications 1 à 3 dans lequel le ou les domaine(s) à activité catalytique a (ont) un pourcentage de similarité compris entre 65 % et 100 % avec la SEQ. ID : n°1.

5. Polypeptide selon l'une des revendications précédentes dans lequel la taille du domaine de liaison au glucane est supérieure à 500 aminoacides.

6. Polypeptide selon la revendication 5. ayant la SEQ.ID : n°2.

7. Polypeptide selon la revendication 6 modifié par substitution, insertion, délétion de séquences d'acides aminés et

moins 90 % de similarité avec les séquences suivantes de la séquence ID

n° 2 :

423 – 439	2120 – 2138
478 – 501	2161 – 2184
519 – 539	2202 – 2214
560 – 571	2243 – 2250
631 – 645	2315 – 2322
1014 – 1021	2689 – 2696

8. Polypeptide selon la revendication 7 dans lequel les acides aminés suivants sont inchangés :

W en positions 425 et 2122,

E en positions 430, 565 et 2127, 2248,

D en positions 487, 489, 527, 638, 2170, 2172, 2210 et 2322,

H en position 637 et 2321,

Q en position 1019 et 2694.

9. Acide nucléique isolé codant une enzyme à activité glycosyltransférase apte à former des dextranes présentant des ramifications  $\alpha 1 \rightarrow 2$  à partir de saccharose, d' $\alpha$ -D-fluoro-glucose, de para-nitrophényl- $\alpha$ -D-glucopyranoside, d' $\alpha$ -D-glucopyranoside- $\alpha$ -D-sorbofuranoside ou de 4-O- $\alpha$ -D-galactopyranosylsucrose, et comprenant au moins une séquence nucléotidique codant un domaine catalytique ayant au moins 50 %, et de préférence au moins 80 % d'identité avec la séquence ID n° 3, et situé en 3' d'une séquence codant un domaine de liaison au glucane.

comportant des séquences présentant au moins 80 % et de préférence au moins 90 % de similarité avec les séquences suivantes de la séquence ID n° 2 :

423 – 439 (SEQ. ID n° 6)	2120 – 2138 (SEQ. ID n° 12)
478 – 501 (SEQ. ID n° 7)	2161 – 2184 (SEQ. ID n° 13)
519 – 539 (SEQ. ID n° 8)	2202 – 2214 (SEQ. ID n° 14)
560 – 571 (SEQ. ID n° 9)	2243 – 2250 (SEQ. ID n° 15)
631 – 645 (SEQ. ID n° 10)	2315 – 2322 (SEQ. ID n° 16)
1014 – 1021 (SEQ. ID n° 11)	2689 – 2696 (SEQ. ID n° 17)

8. Polypeptide selon la revendication 7 dans lequel les acides aminés suivants sont inchangés :

W en positions 425 et 2122,

E en positions 430, 565 et 2127, 2248,

D en positions 487, 489, 527, 638, 2170, 2172, 2210 et 2322,

H en position 637 et 2321,

Q en position 1019 et 2694.

9. Acide nucléique isolé codant une enzyme à activité glycosyltransférase apte à former des dextranes présentant des ramifications  $\alpha 1 \rightarrow 2$  à partir de saccharose, d' $\alpha$ -D-fluoro-glucose, de para-nitrophényl- $\alpha$ -D-glucopyranoside, d' $\alpha$ -D-glucopyranoside- $\alpha$ D-sorbofuranoside ou de 4-O- $\alpha$ -D-galactopyranosylsucrose, et comprenant au moins une séquence nucléotidique codant un domaine catalytique ayant au moins 50 %, et de préférence au moins 80 % d'identité avec la séquence ID n° 3, et situé en 3' d'une séquence codant un domaine de liaison au glucane.

10. Acide nucléique selon la revendication 9 comprenant :

a) deux séquences codant des domaines catalytiques ayant au moins 50 %, et de préférence au moins 80 % d'identité avec la séquence ID n° 3 (1324 – 3922) ;

b) une séquence codant le domaine de liaison au glucane, ce dernier étant situé de préférence entre les deux séquences en a).

11. Acide nucléique isolé selon la revendication 10 présentant au moins 80 % d'identité avec

a) la séquence ID n° 4 ou ;

b) le brin complémentaire de la séquence en a), ou

c) une séquence hybridant a) ou b) en conditions stringentes.

12. Acide nucléique isolé selon la revendication 11 constitué de la séquence ID n° 4 ou son brin complémentaire ou de la séquence déduite de la dégénérescence du code génétique.

13. Acide nucléique isolé selon la revendication 11 comprenant :

a) une séquence ayant au moins 80 % d'identité avec la séquence codant une dextrane saccharase exprimée par le plasmide pCR-T7-dsr D déposé à la CNM le 15 mars 2001 sous le numéro I-2669, ou

b) une séquence complémentaire de la séquence en a).

14. Vecteur d'expression comprenant un acide nucléique selon l'une quelconque des revendications 9 à 13.

15. Vecteur selon la revendication 14 dans lequel l'acide nucléique est sous le contrôle de séquence permettant son expression dans des cellules procaryotes ou eucaryotes.

10. Acide nucléique selon la revendication 9 comprenant :

a) deux séquences codant des domaines catalytiques ayant au moins 50 %, et de préférence au moins 80 % d'identité avec la séquence ID n° 3 ;

b) une séquence codant le domaine de liaison au glucane, ce dernier étant situé de préférence entre les deux séquences en a).

11. Acide nucléique isolé selon la revendication 10 présentant au moins 80 % d'identité avec

a) la séquence ID n° 4 ou ;

b) le brin complémentaire de la séquence en a), ou

c) une séquence hybridant a) ou b) en conditions stringentes.

12. Acide nucléique isolé selon la revendication 11 constitué de la séquence ID n° 4 ou son brin complémentaire ou de la séquence déduite de la dégénérescence du code génétique.

13. Acide nucléique isolé selon la revendication 11 comprenant :

a) une séquence ayant au moins 80 % d'identité avec la séquence codant une dextrane saccharase exprimée par le plasmide pCR-T7-dsr D déposé à la CNCM le 15 mars 2001 sous le numéro I-2649, ou

b) une séquence complémentaire de la séquence en a).

14. Vecteur d'expression comprenant un acide nucléique selon l'une quelconque des revendications 9 à 13.

15. Vecteur selon la revendication 14 dans lequel l'acide nucléique est sous le contrôle de séquence permettant son expression dans des cellules procaryotes ou eucaryotes.

16. Cellule hôte transformée par un acide nucléique selon l'une des revendications 9 à 13 ou un vecteur selon l'une des revendications 14 à 15.

5 17. Cellule transformée selon la revendication 16 choisie dans un groupe comprenant *E. coli*, les leuconostoc, les végétaux, les *Lactococcus* et les *Bacillus* ou les levures ;

10 18 cellule transformée selon la revendication 17 caractérisée en ce qu'il s'agit d'une souche de *E.coli* déposée à la CNCM le... sous le n°...

19. Procédé de production d'une dextrane saccharase apte à former des liaisons  $\alpha 1 \rightarrow 2$  comprenant :

15 a) l'insertion d'un acide nucléique selon l'une des revendications 9 à 13 ou un vecteur selon l'une des revendications 14 à 15 dans une cellule hôte selon la revendication 16;

b) la purification de l'enzyme à partir d'un extrait cellulaire.

20 20. Procédé selon la revendication 19 dans lequel la cellule hôte est un procaryote choisi dans un groupe comprenant *E. coli*, les *Lactococci*, les *Bacillus*, les *Leuconostoc*..

25 21. Procédé selon la revendication 19 dans lequel la cellule hôte est un eucaryote choisi dans un groupe comprenant les levures, les champignons, les végétaux.

30 22. Procédé d'obtention d'une dextrane saccharase apte à former des oligosides ou des dextrans présentant un taux de liaison  $\alpha(1 \rightarrow 2)$  supérieur à 30 % des liaisons totales comprenant une étape de

16. Cellule hôte transformée par un acide nucléique selon l'une des revendications 9 à 13 ou un vecteur selon l'une des revendications 14 à 15.

5 17. Cellule transformée selon la revendication 16 choisie dans un groupe comprenant *E. coli*, les leuconostoc, les végétaux, les *Lactococcus* et les *Bacillus* ou les levures ;

10 18 cellule transformée selon la revendication 17 caractérisée en ce qu'il s'agit d'une souche de *E. coli* déposée à la CNCM le 15 mars 2001 sous le n° 1-2649.

19. Procédé de production d'une dextrane saccharase apte à former des liaisons  $\alpha 1 \rightarrow 2$  comprenant :

15 a) l'insertion d'un acide nucléique selon l'une des revendications 9 à 13 ou un vecteur selon l'une des revendications 14 à 15 dans une cellule hôte selon la revendication 16;

b) la purification de l'enzyme à partir d'un extrait cellulaire.

20 20. Procédé selon la revendication 19 dans lequel la cellule hôte est un procaryote choisi dans un groupe comprenant *E. coli*, les *Lactococci*, les *Bacillus*, les *Leuconostoc*.

25 21. Procédé selon la revendication 19 dans lequel la cellule hôte est un eucaryote choisi dans un groupe comprenant les levures, les champignons, les végétaux.

30 22. Procédé d'obtention d'une dextrane saccharase apte à former des oligosides ou des dextrans présentant un taux de liaison  $\alpha(1 \rightarrow 2)$  supérieur à 30 % des liaisons totales comprenant une étape de



modification de la séquence ID n° 4 par addition, délétion, mutation à partir du moment où :

- le cadre de lecture n'est pas modifié, et
- les acides aminés suivants sont conservés après traduction :

W en positions 425 ou 2122, codé par le triplet TGG en positions 1273 et 6364,

E en positions 430, 565, 2127 et 2248 codés par les triplets GAA en positions 1288, 1693, 6379 et 6742 respectivement,

D en positions 487, 489, 527, 638, 2170 et 2210 codés par les triplets GAT en positions 1459, 1465, 1579, 1912, 6508 et 6628 respectivement,

D en positions 2172 et 2322 codés par les triplets GAT en positions 6514 et 6964,

H en position 637 et 2321, codés respectivement par les triplets CAT en position 1909 et CAC en position 6961,

Q en positions 1019 et 2694 codés respectivement par les triplets CAA (3055) et CAG (8080).

23. Procédé d'obtention d'une glycosyltransférase isolée apte à former des oligosides ou des dextrans présentant un taux de liaison  $\alpha(1 \rightarrow 2)$  supérieur à 30 % comprenant :

- une étape de modification aléatoire de la séquence ID n° 4 et d'établissement d'une banque de variants,

- une étape d'expression de ces séquences modifiées dans une cellule hôte appropriée, un hôte abritant un variant,

- une étape de criblage des hôtes exprimant une enzyme apte à former plus de 30 % de liaison  $\alpha(1 \rightarrow 2)$  sur un substrat approprié,

- une étape d'isolement du ou des gènes améliorés.

modification de la séquence ID n° 4 par addition, délétion, mutation à partir du moment où :

- le cadre de lecture n'est pas modifié, et
- les acides aminés suivants sont conservés après traduction :

5 W en positions 425 ou 2122, codé par le triplet TGG en positions 1273 et 6364,

E en positions 430, 565, 2127 et 2248 codés par les triplets GAA en positions 1288, 1693, 6379 et 6742 respectivement,

10 D en positions 487, 489, 527, 638, 2170 et 2210 codés par les triplets GAT en positions 1459, 1465, 1579, 1912, 6508 et 6628 respectivement,

D en positions 2172 et 2322 codés par les triplets GAT en positions 6514 et 6964,

15 H en position 637 et 2321, codés respectivement par les triplets CAT en position 1909 et CAC en position 6961,

Q en positions 1019 et 2694 codés respectivement par les triplets CAA (3055) et CAG (8080).

20 23. Procédé d'obtention d'une glycosyltransférase isolée apte à former des oligosides ou des dextrans présentant un taux de liaison  $\alpha(1 \rightarrow 2)$  supérieur à 30 % comprenant :

- une étape de modification aléatoire de la séquence ID n° 4 et d'établissement d'une banque de variants;

25 - une étape d'expression de ces séquences modifiées dans une cellule hôte appropriée, un hôte abritant un variant,

- une étape de criblage des hôtes exprimant une enzyme apte à former plus de 30 % de liaison  $\alpha(1 \rightarrow 2)$  sur un substrat approprié,

- une étape d'isolement du ou des gènes améliorés.

30

24. glycosyltransférase apte à former au moins 30 % de liaison  $\alpha(1 \rightarrow 2)$  susceptible d'être obtenue par un procédé selon l'une des revendications 19 à 22.

5 25. Utilisation d'une glycosyltransférase obtenue par un procédé selon l'une des revendications 19 à 22 dans la fabrication d'une composition à effet prébiotique.

10 26. Utilisation d'une glycosyltransférase obtenue par un procédé selon l'une des revendications 19 à 23 dans la fabrication d'une composition pharmaceutique ou cosmétique.

24. glycosyltransférase apte à former au moins 30 % de liaison  $\alpha(1 \rightarrow 2)$  susceptible d'être obtenue par un procédé selon l'une des revendications 19 à 22.

5 25. Utilisation d'une glycosyltransférase obtenue par un procédé selon l'une des revendications 19 à 22 dans la fabrication d'une composition à effet prébiotique.

10 26. Utilisation d'une glycosyltransférase obtenue par un procédé selon l'une des revendications 19 à 23 dans la fabrication d'une composition pharmaceutique ou cosmétique.

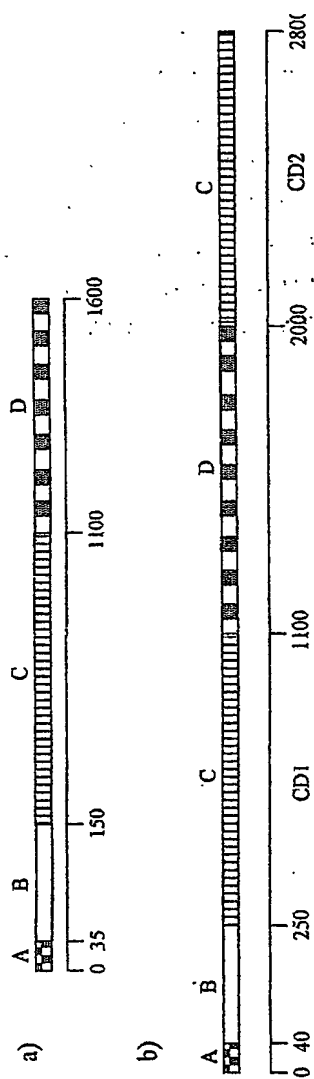


Fig. 1

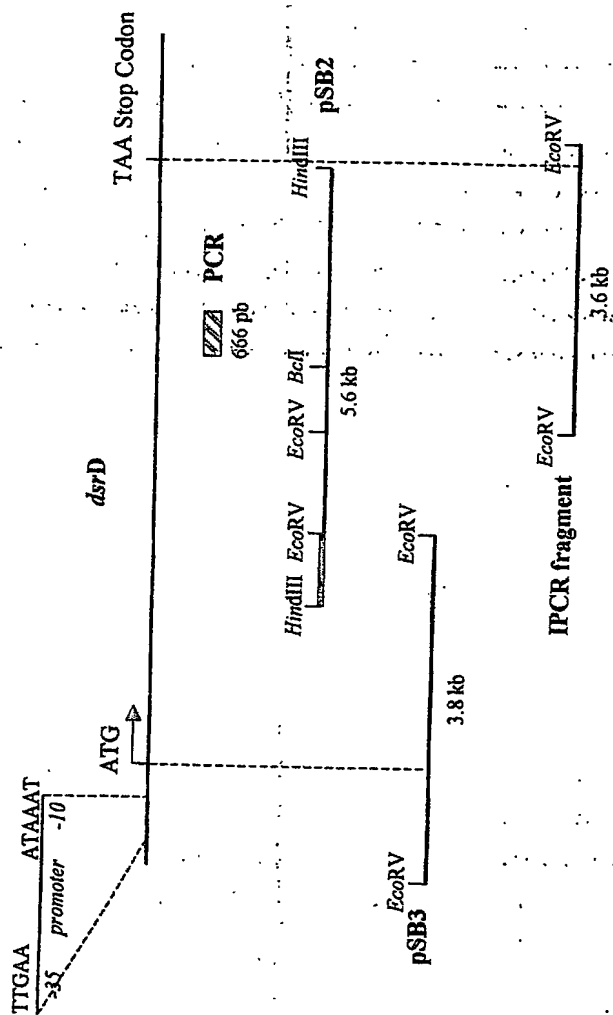


Fig. 2

DSR-D MRDMRVICDRKELKSGKVLVTAG-IFALMMFGVTTASVSA  
DSR-B MFMIKERNVRRKLYKSGKSWVIGLLSLTLMISMTATS-  
DSR-S -MPFTEKVNRRKLYKSGKSWVVG---VCAFALTAS---  
ASR -MKOQETVTRRK-YKSGKVWVAATAFAVLGVSTVTTHA-

Fig. 3

73 AAKVAVATTP-AT  
 86 PVADKTUSA  
 95 PAADKAVDTTSST  
 109 PATDKAVDTTP-IT  
 122 PAADKAVDTTP-IT  
 135 PAADKAVDTTP-IT  
 148 PAANKAVDTTP-AT  
 161 AATDKAV-ATP-AT  
 173 PAADKLANIT--AT  
 185 -----DKAVATTP-AT  
 196 EVANKAA  
 PAADKAVDTTP-AT  
 ← Proposed consensus sequence for S repeat

Fig. 4

Fig. 5

A

GTfB	341	SAWNSDSEK-----PFDDHLQ	402	GGYELLANDVDNSNPVQAEQLN
GTfI	341	PQWNGESEK-----PYDDHLQ	404	GGYELLANDVDNSNPVQAEQLN
GTfS	327	NOWSIASENETVPNQDHMOG	388	AGYELLANDVDNSNPVQAEQLN
dsrS	444	PQWNETSED-----MSNDHLQ	502	GGFELLANDVDNSNPVQAEQLN
dsrA	181	PWNIPSEA-----KGDHHLQ	237	GGFELLANDVDNSNPVQAEQLN
dsrB	426	PQWNMSSED-----PKNDHLQ	484	GGFELLANDVDNSNPVQAEQLN
asr	525	ANWKNQTEDEAF-DGLQWLQ	585	KGEFLLANDIDNSNPVQAEQLN
CD1	423	ANWNIDSES-----KGDHHLQ	478	GGYEMLLANDVDNSNPVQAEQLN
CD2	2120	FIWKNDSYHG--GGDAWFQ	2161	NAFDPELLANDVDNSNPVQAEQLN

B

C

D

E

F

A

GTfB	443	ANFDSIRVDAVDNVDADLLQI	484	HLSILEAWSND	555	YSFIRAHDSVQDLI	928	DWVPDQWY
GTfI	445	ANFDSIRVDAVDNVDADLLQI	486	HVSIVEAWSND	557	YSFIRAHDSVQDLI	932	DWVPDQWY
GTfS	429	ANFDGVRVDAVDNVDADLLQI	470	HLSILEAWSND	540	YVFIRAHDSVQTRI	915	DLVPNQLY
dsrS	543	ANFDGIRVDAVDNVDADLLQI	584	HLSILEDSHND	655	YSFIRAHDSVQTVI	1024	DWVPDQIY
dsrA	278	ANFDGYRVDAVDNVDADLLQI	319	IYQFWKTGEMKI	390	YSFIRAHDSVQTVI	765	DWVPDQIY
dsrB	525	ANFDGIRVDAVDNVDADLLQI	566	HLSILEDSHND	637	YSFIRAHDSVQTVI	1005	DWVPDQIY
asr	626	ANFDGIRVDAVDNVDADLLKI	667	HLSILEDSHND	759	YSFIRAHDSVQTVI	1168	DWVPDQIY
CD1	519	ANFDGYRVDAVDNVDADLLQI	560	HISILEDSHND	631	YAFIRAHDSVQTVI	1014	DWVPDQIY
CD2	2202	ANFDSIRIDAVDFIHNDTIQR	2243	HISLVEAG----	2315	YSIIRAHDKGVQEKV	2689	DVVDNQVY

B

C

D

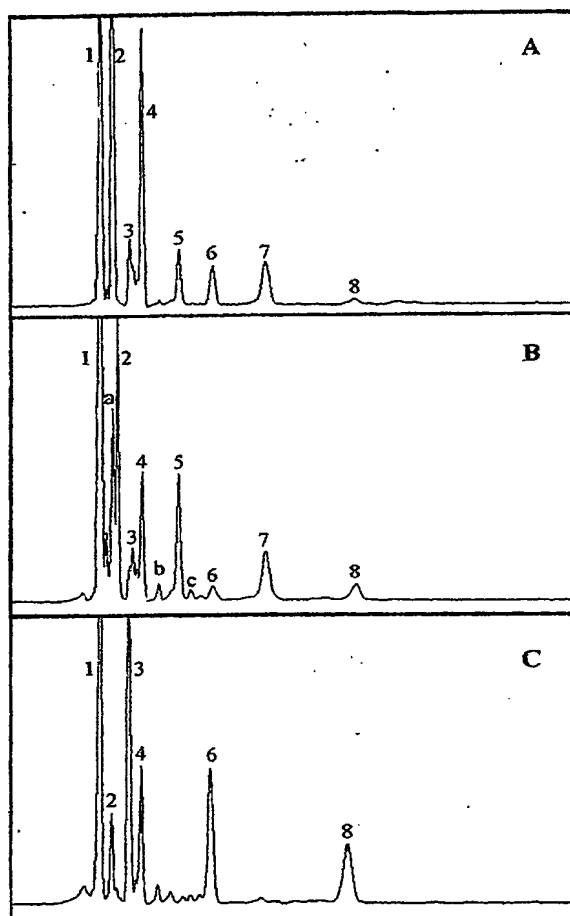
E

F



5/37

**Fig. 6**



## Fig. 7

## (i) CARACTERISTIQUES DE LA SEQUENCE :

(A) LONGUEUR : 856 acides aminés

(B) TYPE : acides aminés

(A) TOPOLOGIE : linéaire

## (ii) TYPE DE MOLECULE : protéine

## (x) DESCRIPTION DE LA SEQUENCE : SEQ ID NO:1:

Leu  
 1980  
 Asp Met Ser Thr Asn Ala Phe Ser Thr Lys Asn Val Ala Phe Asn  
 1985 1990 1995  
 His Asp Ser Ser Ser Phe Asp His Thr Val Asp Gly Phe Leu Thr  
 2000 2005 2010  
 Ala Asp Thr Trp Tyr Arg Pro Lys Ser Ile Leu Ala Asn Gly Thr  
 2015 2020 2025  
 Thr Trp Arg Asp Ser Thr Asp Lys Asp Met Arg Pro Leu Ile Thr  
 2030 2035 2040  
 Val Trp Trp Pro Asn Lys Asn Val Gln Val Asn Tyr Leu Asn Phe  
 2045 2050 2055  
 Met Lys Ala Asn Gly Leu Leu Thr Thr Ala Ala Gln Tyr Thr Leu  
 2060 2065 2070  
 His Ser Asp Gln Tyr Asp Leu Asn Gln Ala Ala Gln Asp Val Gln  
 2075 2080 2085  
 Val Ala Ile Glu Arg Arg Ile Ala Ser Glu His Gly Thr Asp Trp  
 2090 2095 2100  
 Leu Gln Lys Leu Leu Phe Glu Ser Gln Asn Asn Asn Pro Ser Phe  
 2105 2110 2115  
 Val Lys Gln Gln Phe Ile Trp Asn Lys Asp Ser Glu Tyr His Gly  
 2120 2125 2130  
 Gly Gly Asp Ala Trp Phe Gln Gly Gly Tyr Leu Lys Tyr Gly Asn  
 2135 2140 2145

Asn Pro Leu Thr Pro Thr Thr Asn Ser Asp Tyr Arg Gln Pro Gly	2150	2155	2160
Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp Asn Ser Asn	2165	2170	2175
Pro Val Val Gln Ala Glu Asn Leu Asn Trp Leu His Tyr Leu Met	2180	2185	2190
Asn Phe Gly Thr Ile Thr Ala Gly Gln Asp Asp Ala Asn Phe Asp	2195	2200	2205
Ser Ile Arg Ile Asp Ala Val Asp Phe Ile His Asn Asp Thr Ile	2210	2215	2220
Gln Arg Thr Tyr Asp Tyr Leu Arg Asp Ala Tyr Gln Val Gln Gln	2225	2230	2235
Ser Glu Ala Lys Ala Asn Gln His Ile Ser Leu Val Glu Ala Gly	2240	2245	2250
Leu Asp Ala Gly Thr Ser Thr Ile His Asn Asp Ala Leu Ile Glu	2255	2260	2265
Ser Asn Leu Arg Glu Ala Ala Thr Leu Ser Leu Thr Asn Glu Pro	2270	2275	2280
Gly Lys Asn Lys Pro Leu Thr Asn Met Leu Gln Asp Val Asp Gly	2285	2290	2295
Gly Thr Leu Ile Thr Asp His Thr Gln Asn Ser Thr Glu Asn Gln	2300	2305	2310
Ala Thr Pro Asn Tyr Ser Ile Ile His Ala His Asp Lys Gly Val	2315	2320	2325
Gln Glu Lys Val Gly Ala Ala Ile Thr Asp Ala Thr Gly Ala Asp	2330	2335	2340
Trp Thr Asn Phe Thr Asp Glu Gln Leu Lys Ala Gly Leu Glu Leu	2345	2350	2355
Phe Tyr Lys Asp Gln Arg Ala Thr Asn Lys Lys Tyr Asn Ser Tyr	2360	2365	2370
Asn Ile Pro Ser Ile Tyr Ala Leu Met Leu Thr Asn Lys Asp Thr	2375	2380	2385
Val Pro Arg Met Tyr Tyr Gly Asp Met Tyr Gln Asp Asp Gly Gln	2390	2395	2400

Tyr Met Ala Asn Lys Ser Ile Tyr Tyr Asp Ala Leu Val Ser Leu	2405	2410	2415
Met Thr Ala Arg Lys Ser Tyr Val Ser Gly Gly Gln Thr Met Ser	2420	2425	2430
Val Asp Asn His Gly Leu Leu Lys Ser Val Arg Phe Gly Lys Asp	2435	2440	2445
Ala Met Thr Ala Asn Asp Leu Gly Thr Ser Ala Thr Arg Thr Glu	2450	2455	2460
Gly Leu Gly Val Ile Ile Gly Asn Asp Pro Lys Leu Gln Leu Asn	2465	2470	2475
Asp Ser Asp Lys Val Thr Leu Asp Met Gly Ala Ala His Lys Asn	2480	2485	2490
Gln Lys Tyr Arg Ala Val Ile Leu Thr Thr Arg Asp Gly Leu Ala	2495	2500	2505
Thr Phe Asn Ser Asp Gln Ala Pro Thr Ala Trp Thr Asn Asp Gln	2510	2515	2520
Gly Thr Leu Thr Phe Ser Asn Gln Glu Ile Asn Gly Gln Asp Asn	2525	2530	2535
Thr Gln Ile Arg Gly Val Ala Asn Pro Gln Val Ser Gly Tyr Leu	2540	2545	2550
Ala Val Trp Val Pro Val Gly Ala Ser Asp Asn Gln Asp Ala Arg	2555	2560	2565
Thr Ala Ala Thr Thr Thr Glu Asn His Asp Gly Lys Val Leu His	2570	2575	2580
Ser Asn Ala Ala Leu Asp Ser Asn Leu Ile Tyr Glu Gly Phe Ser	2585	2590	2595
Asn Phe Gln Pro Lys Ala Thr Thr His Asp Glu Leu Thr Asn Val	2600	2605	2610
Val Ile Ala Lys Asn Ala Asp Val Phe Asn Asn Trp Gly Ile Thr	2615	2620	2625
Ser Phe Glu Met Ala Pro Gln Tyr Arg Ser Ser Gly Asp His Thr	2630	2635	2640
Phe Leu Asp Ser Thr Ile Asp Asn Gly Tyr Ala Phe Thr Asp Arg	2645	2650	2655

Tyr Asp Leu Gly Phe Asn Thr Pro Thr Lys Tyr Gly Thr Asp Gly	2660	2665	2670
Asp Leu Arg Ala Thr Ile Gln Ala Leu His His Ala Asn Met Gln	2675	2680	2685
Val Met Ala Asp Val Val Asp Asn Gln Val Tyr Asn Leu Pro Gly	2690	2695	2700
Lys Glu Val Val Ser Ala Thr Arg Ala Gly Val Tyr Gly Asn Asp	2705	2710	2715
Asp Ala Thr Gly Phe Gly Thr Gln Leu Tyr Val Thr Asn Ser Val	2720	2725	2730
Gly Gly Gly Gln Tyr Gln Glu Lys Tyr Ala Gly Gln Tyr Leu Glu	2735	2740	2745
Ala Leu Lys Ala Lys Tyr Pro Asp Leu Phe Glu Gly Lys Ala Tyr	2750	2755	2760
Asp Tyr Trp Tyr Lys Asn Tyr Ala Asn Asp Gly Ser Asn Pro Tyr	2765	2770	2775
Tyr Thr Leu Ser His Gly Asp Arg Glu Ser Ile Pro Ala Asp Val	2780	2785	2790
Ala Ile Lys Gln Trp Ser Ala Lys Tyr Met Asn Gly Thr Asn Val	2795	2800	2805
Leu Gly Asn Gly Met Gly Tyr Val Leu Lys Asp Trp His Asn Gly	2810	2815	2820
Gln Tyr Phe Lys Leu Asp Gly Asp Lys Ser Thr Leu Pro Gln Ile	2825	2830	2835

10/37

# Fig. 8

(i) CARACTERISTIQUES DE LA SEQUENCE :

(A) LONGUEUR : 2835 amino acids

(B) TYPE: acide aminé

(A) TOPOLOGIE linéaire

(ii) TYPE DE MOLECULE : protéine

(x) DESCRIPTION DE LA SEQUENCE : SEQ ID NO:2

Met	Arg	Asp	Met	Arg	Val	Ile	Cys	Asp	Arg	Lys	Lys	Leu	Tyr	Lys	1	5	10	15
Ser	Gly	Lys	Val	Leu	Val	Thr	Ala	Gly	Ile	Phe	Ala	Leu	Met	Met	20	25	30	
Phe	Gly	Val	Thr	Thr	Ala	Ser	Val	Ser	Ala	Asn	Thr	Ile	Ala	Val	35	40	45	
Asp	Thr	Asn	His	Ser	Arg	Thr	Ser	Ala	Gln	Ile	Asn	Lys	Ser	Ala	50	55	60	
Val	Asp	Lys	Val	Asn	Asp	Asp	Lys	Thr	Thr	Leu	Gly	Ala	Ala	Lys	65	70	75	
Val	Val	Ala	Val	Ala	Thr	Thr	Pro	Ala	Thr	Pro	Val	Ala	Asp	Lys	80	85	90	
Thr	Val	Ser	Ala	Pro	Ala	Ala	Asp	Lys	Ala	Val	Asp	Thr	Thr	Ser	95	100	105	
Ser	Thr	Thr	Pro	Ala	Thr	Asp	Lys	Ala	Val	Asp	Thr	Thr	Pro	Thr	110	115	120	
Thr	Pro	Ala	Ala	Asp	Lys	Ala	Val	Asp	Thr	Thr	Pro	Thr	Thr	Pro	125	130	135	
Ala	Ala	Asp	Lys	Ala	Val	Asp	Thr	Thr	Pro	Thr	Thr	Pro	Ala	Ala	140	145	150	
Asn	Lys	Ala	Val	Asp	Thr	Thr	Pro	Ala	Thr	Ala	Ala	Thr	Asp	Lys	155	160	165	
Ala	Val	Ala	Thr	Pro	Ala	Thr	Pro	Ala	Ala	Asp	Lys	Leu	Ala	Asn				

Thr Thr Pro Ala Thr Asp Lys Ala Val Ala Thr Thr Pro Ala Thr	185	190	195
Pro Val Ala Asn Lys Ala Ala Asp Thr Ser Ser Ile His Asp Gln	200	205	210
Pro Leu Asp Thr Asn Val Pro Thr Asp Lys Ser Ala Asn Leu Val	215	220	225
Ser Thr Thr Gln Lys Ser Thr Asp Asn Gln Gln Val Lys Ser Thr	230	235	240
Glu Thr Ser His Leu Gln Glu Ile Asn Gly Lys Thr Tyr Phe Leu	245	250	255
Asp Asp Asn Gly Gln Val Lys Lys Asn Phe Thr Ala Ile Ile Asp	260	265	270
Gly Lys Val Leu Tyr Phe Asp Lys Thr Ser Gly Glu Leu Thr Ala	275	280	285
Asn Ala Pro Gln Val Thr Lys Gly Leu Val Asn Ile Asp Asn Ala	290	295	300
His Asn Ala Ala His Asp Leu Thr Ala Asp Asn Phe Thr Asn Val	305	310	315
Asp Gly Tyr Leu Thr Ala Asn Ser Trp Tyr Arg Pro Lys Asp Ile	320	325	330
Leu Lys Asn Gly Thr Thr Trp Thr Pro Thr Thr Ala Glu Asp Phe	335	340	345
Arg Pro Leu Leu Met Ser Trp Trp Pro Asp Lys Asn Thr Gln Val	350	355	360
Ala Tyr Leu Gln Tyr Met Gln Ser Val Gly Met Leu Pro Asp Asp	365	370	375
Val Lys Val Ser Asn Asp Asp Asn Met Ser Thr Leu Thr Asp Ala	380	385	390
Ala Met Thr Val Gln Lys Asn Ile Glu Ser Arg Ile Gly Val Ser	395	400	405
Gly Lys Thr Asp Trp Leu Lys Gln Asp Met Asn Lys Leu Ile Asp	410	415	420
Ser Gln Ala Asn Trp Asn Ile Asp Ser Glu Ser Lys Gly Asn Asp	425	430	435

His	Leu	Gln	Gly	Gly	Ala	Leu	Leu	Tyr	Val	Asn	Asp	Asp	Lys	Thr	440	445	450
Pro	Asn	Ala	Asn	Ser	Asp	Tyr	Arg	Leu	Leu	Asn	Arg	Thr	Pro	Thr	455	460	465
Asn	Gln	Thr	Gly	Gln	Ile	Thr	Asp	Pro	Ser	Lys	Gln	Gly	Gly	Tyr	470	475	480
Glu	Met	Leu	Leu	Ala	Asn	Asp	Val	Asp	Asn	Ser	Asn	Pro	Val	Val	485	490	495
Gln	Ala	Glu	Gln	Leu	Asn	Trp	Leu	His	Tyr	Met	Met	Asn	Ile	Gly	500	505	510
Thr	Ile	Ala	Gln	Asn	Asp	Pro	Thr	Ala	Asn	Phe	Asp	Gly	Tyr	Arg	515	520	525
Val	Asp	Ala	Val	Asp	Asn	Val	Asp	Ala	Asp	Leu	Leu	Gln	Ile	Ala	530	535	540
Gly	Asp	Tyr	Phe	Lys	Ala	Ala	Tyr	Gly	Thr	Gly	Lys	Thr	Glu	Ala	545	550	555
Asn	Ala	Asn	Asn	His	Ile	Ser	Ile	Leu	Glu	Asp	Trp	Asp	Asn	Asn	560	565	570
Asp	Ser	Ala	Tyr	Ile	Lys	Ala	His	Gly	Asn	Asn	Gln	Leu	Thr	Met	575	580	585
Asp	Phe	Pro	Ala	His	Leu	Ala	Leu	Lys	Tyr	Ala	Leu	Asn	Met	Pro	590	595	600
Leu	Ala	Ala	Gln	Ser	Gly	Leu	Glu	Pro	Leu	Ile	Asn	Thr	Ser	Leu	605	610	615
Val	Lys	Arg	Gly	Lys	Asp	Ala	Thr	Glu	Asn	Glu	Ala	Gln	Pro	Asn	620	625	630
Tyr	Ala	Phe	Ile	Arg	Ala	His	Asp	Ser	Glu	Val	Gln	Thr	Val	Ile	635	640	645
Ala	Gln	Ile	Ile	Lys	Asp	Lys	Ile	Asn	Thr	Lys	Ser	Asp	Gly	Leu	650	655	660
Thr	Val	Thr	Pro	Asp	Glu	Ile	Lys	Gln	Ala	Phe	Thr	Ile	Tyr	Asn	665	670	675
Ala	Asp	Glu	Leu	Lys	Ala	Asp	Lys	Glu	Tyr	Thr	Ala	Tyr	Asn	Ile	680	685	690



Pro	Ala	Ser	Tyr	Ala	Val	Leu	Leu	Thr	Asn	Lys	Asp	Thr	Val	Pro	695	700	705
Arg	Val	Tyr	Tyr	Gly	Asp	Leu	Phe	Ser	Asp	Asp	Gly	Gln	Tyr	Met	710	715	720
Ser	Gln	Lys	Ser	Pro	Tyr	Tyr	Asp	Ala	Ile	Thr	Ser	Leu	Leu	Lys	725	730	735
Ser	Arg	Ile	Lys	Tyr	Val	Ala	Gly	Gly	Gln	Ser	Met	Asn	Met	Thr	740	745	750
Tyr	Leu	His	Glu	Cys	Phe	Asp	Pro	Ala	Lys	Asn	Glu	Thr	Lys	Pro	755	760	765
Gln	Gly	Val	Leu	Thr	Ser	Val	Arg	Tyr	Gly	Lys	Gly	Ala	Met	Thr	770	775	780
Ala	Asp	Asp	Leu	Gly	Asn	Ser	Asp	Thr	Arg	Gln	Gln	Gly	Ile	Gly	785	790	795
Leu	Val	Ile	Asn	Asn	Lys	Pro	Phe	Leu	Asn	Leu	Asn	Asp	Asp	Glu	800	805	810
Gln	Ile	Val	Leu	Asn	Met	Gly	Ala	Ala	His	Lys	Asn	Gln	Ala	Tyr	815	820	825
Arg	Pro	Leu	Met	Leu	Thr	Thr	Lys	Ser	Gly	Leu	Gln	Ile	Tyr	Asp	830	835	840
Lys	Asp	Ala	Gly	Ala	Pro	Val	Val	Tyr	Thr	Asn	Asp	Ala	Gly	Gln	845	850	855
Leu	Ile	Phe	Lys	Ser	Asp	Met	Val	Tyr	Gly	Val	Ser	Asn	Pro	Gln	860	865	870
Val	Ser	Gly	Tyr	Phe	Ala	Ala	Trp	Val	Pro	Val	Gly	Ala	Ser	Asp	875	880	885
Ser	Gln	Asp	Ala	Arg	Thr	Gln	Ser	Ser	Gln	Ser	Glu	Thr	Lys	Asp	890	895	900
Gly	Asp	Val	Tyr	His	Ser	Asn	Ala	Ala	Leu	Asp	Ser	Asn	Val	Ile	905	910	915
Tyr	Glu	Gly	Phe	Ser	Asn	Phe	Gln	Ala	Met	Pro	Glu	Lys	Asn	Asp	920	925	930
Asp	Phe	Thr	Asn	Val	Lys	Ile	Ala	Gln	Asn	Ala	Lys	Leu	Phe	Lys	935	940	945

Asp Leu Gly Ile Thr Ser Phe Glu Leu Ala Pro Gln Tyr Arg Ser	950	955	960
Ser Thr Asp Asn Ser Phe Leu Asp Ser Val Ile Gln Asn Gly Tyr	965	970	975
Ala Phe Thr Asp Arg Tyr Asp Val Gly Tyr Asn Thr Pro Thr Lys	980	985	990
Tyr Gly Thr Val Asp Gln Leu Leu Asp Ser Leu Arg Ala Leu His	995	1000	1005
Ala Gln Gly Ile Gln Ala Ile Asn Asp Trp Val Pro Asp Gln Ile	1010	1015	1020
Tyr Asn Leu Pro Gly Glu Gln Ile Val Thr Ala Val Arg Thr Asn	1025	1030	1035
Gly Ser Gly Lys Tyr Asp Tyr Asp Ser Val Ile Asn Asn Thr Leu	1040	1045	1050
Tyr Asp Ser Arg Thr Val Gly Gly Gly Glu Tyr Gln Glu Lys Phe	1055	1060	1065
Gly Gly Leu Phe Leu Asp Gln Leu Lys Lys Asp Tyr Pro Ser Leu	1070	1075	1080
Phe Glu Thr Lys Gln Ile Ser Thr Asn Gln Pro Met Asn Pro Asp	1085	1090	1095
Val Lys Ile Lys Glu Trp Ser Ala Lys Tyr Phe Asn Gly Ser Asn	1100	1105	1110
Ile Gln Gly Arg Gly Ala Trp Tyr Val Leu Lys Asp Trp Ala Thr	1115	1120	1125
Asn Gln Tyr Phe Asn Val Ser Ser Asp Asn Gly Phe Leu Pro Lys	1130	1135	1140
Gln Leu Leu Gly Glu Lys Thr Ser Thr Gly Phe Ile Thr Glu Asn	1145	1150	1155
Gly Lys Thr Ser Phe Tyr Ser Thr Ser Gly Tyr Gln Ala Lys Asp	1160	1165	1170
Thr Phe Ile Gln Asp Gly Thr Asn Trp Tyr Tyr Phe Asp Asn Ala	1175	1180	1185
Gly Tyr Met Leu Thr Gly Lys Gln Asn Ile His Asp Lys Asn Tyr	1190	1195	1200

Tyr Phe Leu Pro Asn Gly Val Glu Leu Gln Asp Ala Tyr Leu Phe  
 1205 1210 1215  
 Asp Gly Asn Gln Glu Phe Tyr Tyr Asn Lys Ala Gly Glu Gln Val  
 1220 1225 1230  
 Met Asn Gln Tyr Tyr Gln Asp Ser Gln Asn Gln Trp His Tyr Phe  
 1235 1240 1245  
 Phe Glu Asn Gly Arg Met Ala Ile Gly Leu Thr Glu Val Pro Asn  
 1250 1255 1260  
 Ala Asp Gly Thr His Val Thr Gln Tyr Phe Asp Ala Asn Gly Val  
 1265 1270 1275  
 Gln Ile Lys Gly Thr Ala Ile Lys Asp Gln Asn Asn Gln Leu Arg  
 1280 1285 1290  
 Tyr Phe Asp Glu Ala Thr Gly Asn Met Val Val Asn Ser Trp Gly  
 1295 1300 1305  
 Gln Leu Ala Asp Lys Ser Trp Leu Tyr Leu Asn Ala Gln Gly Val  
 1310 1315 1320  
 Ala Val Thr Gly Asn Gln Lys Ile Asp Gly Glu Glu Tyr Tyr Phe  
 1325 1330 1335  
 Asn Ala Asp Gly Lys Gln Val Lys Gly Asn Ala Ile Ile Asp Asn  
 1340 1345 1350  
 Asn Gly Asp Gln Arg Tyr Tyr Asp Gly Asp Lys Gly Val Met Val  
 1355 1360 1365  
 Val Asn Ser Trp Gly Glu Leu Pro Asp Gly Ser Trp Leu Tyr Leu  
 1370 1375 1380  
 Asn Asp Lys Gly Ile Ala Val Thr Gly Arg Gln Val Ile Asn Asn  
 1385 1390 1395  
 Gln Val Asn Phe Phe Gly Asn Asp Gly Lys Gln Ile Lys Asp Ala  
 1400 1405 1410  
 Phe Lys Leu Leu Ser Asp Gly Ser Trp Val Tyr Leu Asp Asp Lys  
 1415 1420 1425  
 Gly Leu Ile Thr Thr Gly Ala Lys Val Ile Asn Gly Leu Asn Met  
 1430 1435 1440  
 Phe Phe Asp Lys Asp Gly His Gln Ile Lys Gly Asp Ala Ser Thr  
 1445 1450 1455

Asp Ala Asn Gly Lys Arg His Tyr Tyr Asp Lys Asn Asp Gly His	1460	1465	1470
Leu Val Thr Asn Ser Trp Gly Glu Leu Pro Asp Gly Ser Trp Leu	1475	1480	1485
Tyr Leu Glu Glu Gln Gly Asp Ala Val Thr Gly Gln Arg Val Ile	1490	1495	1500
Asp Gly Lys Thr Arg Tyr Phe Asp Glu Asp Gly Lys Gln Ile Lys	1505	1510	1515
Asn Ser Leu Lys Thr Leu Ala Asn Gly Asp Lys Ile Tyr Leu Asp	1520	1525	1530
Gly Asp Gly Val Ala Ala Thr Gly Leu Gln His Val Gly Asp Lys	1535	1540	1545
Ile Met Tyr Phe Asp Glu Asp Gly Lys Gln Val Val Gly Lys Phe	1550	1555	1560
Val Ser Ala Lys Asp Gly Ser Trp Tyr Tyr Leu Asn Gln Asp Gly	1565	1570	1575
Val Ala Ala Val Gly Pro Ser Ser Ile Asn Gly Gln Ser Leu Tyr	1580	1585	1590
Phe Asp Gln Asp Gly Lys Gln Val Lys Tyr Asn Glu Val Arg Asn	1595	1600	1605
Ser Asp Gly Thr Thr Asn Tyr Tyr Thr Gly Leu Thr Gly Glu Lys	1610	1615	1620
Leu Thr Gln Asp Phe Gly Glu Leu Pro Asp Gly Ser Trp Ile Tyr	1625	1630	1635
Leu Asp Ala Gln Gly His Thr Val Thr Gly Ala Gln Ile Ile Asn	1640	1645	1650
Gly Gln Asn Leu Tyr Phe Lys Ala Asp Gly Gln Gln Val Lys Gly	1655	1660	1665
His Ala Tyr Thr Asp Gln Leu Gly His Met Arg Phe Tyr Asp Pro	1670	1675	1680
Asp Ser Gly Asp Met Leu Ser Asn Arg Phe Glu Gln Ile Thr Pro	1685	1690	1695
Gly Val Trp Ala Tyr Phe Gly Ala Asp Gly Val Ala Ile Thr Gly	1700	1705	1710

Gln His Asp Ile Asn Gly Gln Lys Leu Phe Phe Asp Glu Thr Gly  
 1715 1720 1725  
 Tyr Gln Val Lys Gly Ser Gln Arg Thr Ile Asp Gly Thr Leu Tyr  
 1730 1735 1740  
 Ser Phe Asp Ser Gln Thr Gly Asn Gln Lys Arg Val Gln Thr Thr  
 1745 1750 1755  
 Leu Leu Pro Gln Ala Gly His Tyr Ile Thr Lys Asn Gly Asn Asp  
 1760 1765 1770  
 Trp Gln Tyr Asp Thr Asn Gly Glu Leu Ala Lys Gly Leu Arg Gln  
 1775 1780 1785  
 Asp Ser Asn Gly Lys Leu Arg Tyr Phe Asp Leu Thr Thr Gly Ile  
 1790 1795 1800  
 Gln Ala Lys Gly Gln Phe Val Thr Ile Gly Gln Glu Thr Tyr Tyr  
 1805 1810 1815  
 Phe Ser Lys Asp His Gly Asp Ala Gln Leu Leu Pro Met Val Thr  
 1820 1825 1830  
 Glu Gly His Tyr Gly Thr Ile Thr Leu Lys Gln Gly Gln Asp Thr  
 1835 1840 1845  
 Lys Thr Ala Trp Val Tyr Arg Asp Gln Asn Asn Thr Ile Leu Lys  
 1850 1855 1860  
 Gly Leu Gln Asn Ile Asn Gly Thr Leu Gln Phe Phe Asp Pro Tyr  
 1865 1870 1875  
 Thr Gly Glu Gln Leu Lys Gly Gly Val Ala Lys Tyr Asp Asp Lys  
 1880 1885 1890  
 Leu Phe Tyr Phe Glu Ser Gly Lys Gly Asn Leu Val Ser Thr Val  
 1895 1900 1905  
 Ala Gly Asp Tyr Gln Asp Gly His Tyr Ile Ser Gln Asp Gly Gln  
 1910 1915 1920  
 Thr Arg Tyr Ala Asp Lys Gln Asn Gln Leu Val Lys Gly Leu Val  
 1925 1930 1935  
 Thr Val Asn Gly Ala Leu Gln Tyr Phe Asp Asn Ala Thr Gly Asn  
 1940 1945 1950  
 Gln Ile Lys Asn Gln Gln Val Ile Val Asp Gly Lys Thr Tyr Tyr  
 1955 1960 1965

Phe Asp Asp Lys Gly Asn Gly Glu Tyr Leu Phe Thr Asn Thr Leu	1970	1975	1980
Asp Met Ser Thr Asn Ala Phe Ser Thr Lys Asn Val Ala Phe Asn	1985	1990	1995
His Asp Ser Ser Ser Phe Asp His Thr Val Asp Gly Phe Leu Thr	2000	2005	2010
Ala Asp Thr Trp Tyr Arg Pro Lys Ser Ile Leu Ala Asn Gly Thr	2015	2020	2025
Thr Trp Arg Asp Ser Thr Asp Lys Asp Met Arg Pro Leu Ile Thr	2030	2035	2040
Val Trp Trp Pro Asn Lys Asn Val Gln Val Asn Tyr Leu Asn Phe	2045	2050	2055
Met Lys Ala Asn Gly Leu Leu Thr Thr Ala Ala Gln Tyr Thr Leu	2060	2065	2070
His Ser Asp Gln Tyr Asp Leu Asn Gln Ala Ala Gln Asp Val Gln	2075	2080	2085
Val Ala Ile Glu Arg Arg Ile Ala Ser Glu His Gly Thr Asp Trp	2090	2095	2100
Leu Gln Lys Leu Leu Phe Glu Ser Gln Asn Asn Asn Pro Ser Phe	2105	2110	2115
Val Lys Gln Gln Phe Ile Trp Asn Lys Asp Ser Glu Tyr His Gly	2120	2125	2130
Gly Gly Asp Ala Trp Phe Gln Gly Gly Tyr Leu Lys Tyr Gly Asn	2135	2140	2145
Asn Pro Leu Thr Pro Thr Thr Asn Ser Asp Tyr Arg Gln Pro Gly	2150	2155	2160
Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp Asn Ser Asn	2165	2170	2175
Pro Val Val Gln Ala Glu Asn Leu Asn Trp Leu His Tyr Leu Met	2180	2185	2190
Asn Phe Gly Thr Ile Thr Ala Gly Gln Asp Asp Ala Asn Phe Asp	2195	2200	2205
Ser Ile Arg Ile Asp Ala Val Asp Phe Ile His Asn Asp Thr Ile	2210	2215	2220

19/37

Gln Arg Thr Tyr Asp Tyr Leu Arg Asp Ala Tyr Gln Val Gln Gln  
2225 2230 2235

Ser Glu Ala Lys Ala Asn Gln His Ile Ser Leu Val Glu Ala Gly  
2240 2245 2250

Leu Asp Ala Gly Thr Ser Thr Ile His Asn Asp Ala Leu Ile Glu  
2255 2260 2265

Ser Asn Leu Arg Glu Ala Ala Thr Leu Ser Leu Thr Asn Glu Pro  
2270 2275 2280

Gly Lys Asn Lys Pro Leu Thr Asn Met Leu Gln Asp Val Asp Gly  
2285 2290 2295

Gly Thr Leu Ile Thr Asp His Thr Gln Asn Ser Thr Glu Asn Gln  
2300 2305 2310

Ala Thr Pro Asn Tyr Ser Ile Ile His Ala His Asp Lys Gly Val  
2315 2320 2325

Gln Glu Lys Val Gly Ala Ala Ile Thr Asp Ala Thr Gly Ala Asp  
2330 2335 2340

Trp Thr Asn Phe Thr Asp Glu Gln Leu Lys Ala Gly Leu Glu Leu  
2345 2350 2355

Phe Tyr Lys Asp Gln Arg Ala Thr Asn Lys Lys Tyr Asn Ser Tyr  
2360 2365 2370

Asn Ile Pro Ser Ile Tyr Ala Leu Met Leu Thr Asn Lys Asp Thr  
2375 2380 2385

Val Pro Arg Met Tyr Tyr Gly Asp Met Tyr Gln Asp Asp Gly Gln  
2390 2395 2400

Tyr Met Ala Asn Lys Ser Ile Tyr Tyr Asp Ala Leu Val Ser Leu  
2405 2410 2415

Met Thr Ala Arg Lys Ser Tyr Val Ser Gly Gly Gln Thr Met Ser  
2420 2425 2430

Val Asp Asn His Gly Leu Leu Lys Ser Val Arg Phe Gly Lys Asp  
2435 2440 2445

Ala Met Thr Ala Asn Asp Leu Gly Thr Ser Ala Thr Arg Thr Glu  
2450 2455 2460

Gly Leu Gly Val Ile Ile Gly Asn Asp Pro Lys Leu Gln Leu Asn  
2465 2470 2475 -

20/37

Asp Ser Asp Lys Val Thr Leu Asp Met Gly Ala Ala His Lys Asn	2480	2485	2490
Gln Lys Tyr Arg Ala Val Ile Leu Thr Thr Arg Asp Gly Leu Ala	2495	2500	2505
Thr Phe Asn Ser Asp Gln Ala Pro Thr Ala Trp Thr Asn Asp Gln	2510	2515	2520
Gly Thr Leu Thr Phe Ser Asn Gln Glu Ile Asn Gly Gln Asp Asn	2525	2530	2535
Thr Gln Ile Arg Gly Val Ala Asn Pro Gln Val Ser Gly Tyr Leu	2540	2545	2550
Ala Val Trp Val Pro Val Gly Ala Ser Asp Asn Gln Asp Ala Arg	2555	2560	2565
Thr Ala Ala Thr Thr Thr Glu Asn His Asp Gly Lys Val Leu His	2570	2575	2580
Ser Asn Ala Ala Leu Asp Ser Asn Leu Ile Tyr Glu Gly Phe Ser	2585	2590	2595
Asn Phe Gln Pro Lys Ala Thr Thr His Asp Glu Leu Thr Asn Val	2600	2605	2610
Val Ile Ala Lys Asn Ala Asp Val Phe Asn Asn Trp Gly Ile Thr	2615	2620	2625
Ser Phe Glu Met Ala Pro Gln Tyr Arg Ser Ser Gly Asp His Thr	2630	2635	2640
Phe Leu Asp Ser Thr Ile Asp Asn Gly Tyr Ala Phe Thr Asp Arg	2645	2650	2655
Tyr Asp Leu Gly Phe Asn Thr Pro Thr Lys Tyr Gly Thr Asp Gly	2660	2665	2670
Asp Leu Arg Ala Thr Ile Gln Ala Leu His His Ala Asn Met Gln	2675	2680	2685
Val Met Ala Asp Val Val Asp Asn Gln Val Tyr Asn Leu Pro Gly	2690	2695	2700
Lys Glu Val Val Ser Ala Thr Arg Ala Gly Val Tyr Gly Asn Asp	2705	2710	2715
Asp Ala Thr Gly Phe Gly Thr Gln Leu Tyr Val Thr Asn Ser Val	2720	2725	2730



21/37

Gly	Gly	Gly	Gln	Tyr	Gln	Glu	Lys	Tyr	Ala	Gly	Gln	Tyr	Leu	Glu
					2735					2740				2745
Ala	Leu	Lys	Ala	Lys	Tyr	Pro	Asp	Leu	Phe	Glu	Gly	Lys	Ala	Tyr
					2750					2755				2760
Asp	Tyr	Trp	Tyr	Lys	Asn	Tyr	Ala	Asn	Asp	Gly	Ser	Asn	Pro	Tyr
					2765					2770				2775
Tyr	Thr	Leu	Ser	His	Gly	Asp	Arg	Glu	Ser	Ile	Pro	Ala	Asp	Val
					2780					2785				2790
Ala	Ile	Lys	Gln	Trp	Ser	Ala	Lys	Tyr	Met	Asn	Gly	Thr	Asn	Val
					2795					2800				2805
Leu	Gly	Asn	Gly	Met	Gly	Tyr	Val	Leu	Lys	Asp	Trp	His	Asn	Gly
					2810					2815				2820
Gln	Tyr	Phe	Lys	Leu	Asp	Gly	Asp	Lys	Ser	Thr	Leu	Pro	Gln	Ile
					2825					2830				2835

22/37

Fig. 9

		TTA	5940
		L	1980
5941	GATATGTCTACTAATGCTTTTCTACCAAAATGTTGCATTCAATCATGACAGTAGCAGT		6000
1981	D M S T N A F S T K N V A F N H D S S S		2000
6001	TTCGACCATACTGTTGATGGCTTCTTGACGGCAGATACTTGGTATCGACCAAAGTCAATT		6060
2001	F D H T V D G F L T A D T W Y R P K S I		2020
6061	TTGGCTAACGGGACAACCTGGCGTGATTGACTGATAAGGATATGCGACCATTAATCACT		6120
2021	L A N G T T W R D S T D K D M R P L I T		2040
6121	GTTTGGTGGCCAAATAAGAATGTTCAAGTCAACTACCTCAACTTCATGAAAGCAAATGGC		6180
2041	V W W P N K N V Q V N Y L N F M K A N G		2060
6181	TTGTTGACAACAGCAGCACAATACACACTACATTGAGATCAATATGATTGAACCAAGCT		6240
2061	L L T T A A Q Y T L H S D Q Y D L N Q A		2080
6241	GCACAAGATGTTCAAGTGGCCATTGAAAGGCGCATTGCGTCAGAGCATGGCACAGACTGG		6300
2081	A Q D V Q V A I E R R I A S E H G T D W		2100
6301	TTACAGAAATTGTTGTTGAATCACAAATAATAACCCATCATTTGTGAAGCAACAATTC		6360
2101	L Q K L L F E S Q N N N P S F V K Q Q F		2120
6361	ATTTGGAACAAGGATTCTGAATATCATGGTGGTGGTGATGCTTGGTTCCAAGGTGGTTAT		6420
2121	I W N K D S E Y H G G G D A W F Q G G Y		2140
6421	CTGAAGTATGGCAATAACCCACTCACACCAACAACTAATTCTGATTATCGTCAACCTGGT		6480
2141	L K Y G N N P L T P T T N S D Y R Q P G		2160
6481	AATGCATTTGATTCTTGCTAGCCAACGACGTGGATAATTCTAATCCTGTTGTGCAAGCT		6540
2161	N A F D F L L A N D V D N S N P V V Q A		2180
6541	GAAACTTAAACTGGTTACATTACTTAATGAAGTGGCACCATCACTGCGGGTCAAGAT		6600
2181	E N L N W L H Y L M N F G T I T A G Q D		2200
6601	GACGCTAATTTGATAGTATTCGTATTGACGCTGTCGACTTTATTCATAATGATACAATC		6660
2201	D A N F D S I R I D A V D F I H N D T I		2220
6661	CAACGTACTTATGATTATCTTCGTGATGCTTATCAAGTGCAACAAAGTGAAGCCAAAGCA		6720
2221	Q R T Y D Y L R D A Y Q V Q Q S E A K A		2240
6721	AACCAGCACATTTCAATGGTTGAAGCTGGCTTAGACGCAGGTACATCAACGATTTCATAAT		6780
2241	N Q H I S L V E A G L D A G T S T I H N		2260
6781	GATGCGTTAATTGAGTCAAACCTCCGTGAAGCAGCGACATTGTCGTTAAACAAATGAACCT		6840
2261	D A L I E S N L R E A A T L S L T N E P		2280
6841	GGTAAAAATAAACCATGACGAATATGCTACAAGACGTTGACGGCGGTACGCTTATCACC		6900
2281	G K N K P L T N M L Q D V D G G T L I T		2300

6901	GACCATACGCAGAATAGTACAGAAAATCAGGCGACACCAAACCTATTCAATTATTCACGCG	6960
2301	D H T Q N S T E N Q A T P N Y S I I H A	2320
6961	CACGATAAAGGTGTGCAAGAAAAAGTAGGTGCAGCCATTACTGATGCTACTGGTGCTGAT	7020
2321	H D K G V Q E K V G A A I T D A T G A D	2340
7021	TGGACGAACCTTTACAGATGAACAGTTAAAAGCCGGATTAGAGCTATTCTATAAGGATCAG	7080
2341	W T N F T D E Q L K A G L E L F Y K D Q	2360
7081	CGCGCAACAAACAAAAAGTATAATAGTTATAACATACCAAGTATTTATGCCCTGATGTTG	7140
2361	R A T N K K Y N S Y N I P S I Y A L M L	2380
7141	ACAAACAAAGATACTGTTCTCGTATGTATTATGGGGATATGTATCAAGATGACGGACAG	7200
2381	T N K D T V P R M Y Y G D M Y Q D D G Q	2400
7201	TATATGGCAACAAAGAGTATCTACTATGATGCCTTAGTGTCAATTAATGACGGCTCGTAAA	7260
2401	Y M A N K S I Y Y D A L V S L M T A R K	2420
7261	AGCTATGTCAGCGGTGGTCAAACCTATGAGTGTGACAATCATGGTTTGTGAAGAGTGTC	7320
2421	S Y V S G G Q T M S V D N H G L L K S V	2440
7321	CGTTTTGGAAAAGATGCGATGACAGCTAATGATTTAGGTACATCAGCTACGCGTACTGAG	7380
2441	R F G K D A M T A N D L G T S A T R T E	2460
7381	GGTCTTGGTGTCAATTATTGGTAATGATCCAAAGTTGCAACTTAATGATTCGGATAAAGTG	7440
2461	G L G V I I G N D P K L Q L N D S D K V	2480
7441	ACACTGGATAIGGGTGCAGCACATAAAAATCAAAGTATCGCGCAGTTATCTTAACAACA	7500
2481	T L D M G A A H K N Q K Y R A V I L T T	2500
7501	CGTGTATGGTTTGGCAACCTTTAATTCAGATCAAGCACCAACAGCTTGGACAAACGATCAA	7560
2501	R D G L A T F N S D Q A P T A W T N D Q	2520
7561	CGAACGTTAACATTCTCAAATCAAGAGATTAACGGGCAAGACAATACACAAATTCGTGGT	7620
2521	G T L T F S N Q E I N G Q D N T Q I R G	2540
7621	GTTGCTAATCCGCAAGTTTCTGGTTATCTAGCTGTTTGGGTGCCTGTGGGTGCATCAGAC	7680
2541	V A N P Q V S G Y L A V W V P V G A S D	2560
7681	AATCAAGATGCCCGTACAGCAGCAACGACAACAGAAAATCATGATGGTAAAGTATTACAC	7740
2561	N Q D A R T A A T T T E N H D G K V L H	2580
7741	TCGAATGCGGCATTAGATTCTAACCTTATTATGAAGGTTTCTCTAACTTCCAACTAAG	7800
2581	S N A A L D S N L I Y E G F S N F Q P K	2600
7801	GCAACAACGCATGATGAACCTACGAACGTTGTAATTGCTAAAATGCCGATGTCTTCAAT	7860
2601	A T T H D E L T N V V I A K N A D V F N	2620
7861	AATTGGGGTATTACGAGTTTGAATGGCACCACAGTACCGTTCAAGTGGGGACCATACA	7920
2621	N W G I T S F E M A P Q Y R S S G D H T	2640
7921	TTCTTGGATTCAACGATTGATAATGGTTATGCCTTCACTGATCGCTATGACTTAGGTTTC	7980
2641	F L D S T I D N G Y A F T D R Y D L G F	2660
7981	AATACACCAACAAAGTATGGCACTGATGGTGATTGCGTGCAACGATTCAAGCGCTACAT	8040
2661	N T P T K Y G T D G D L R A T I Q A L H	2680
8041	CATGCTAATATGCAAGTTATGGCTGACGTTGTTGATAACCAGGTCTATAACTTACCTGGT	8100
2681	H A N M Q V M A D V V D N Q V Y N L P G	2700

8101	AAAGAAGTTGTTTCAGCAACACGAGCAGGTGTTTATGGTAATGACGACGCCACGGGCTTT	8160
2701	K E V V S A T R A G V Y G N D D A T G F	2720
8161	GGAACGCAACTCTATGTGACTAACTCCGTTGGTGGTGGTCAATACCAAGAGAAATATGCT	8220
2721	G T Q L Y V T N S V G G G Q Y Q E K Y A	2740
8221	GGACAATACTTAGAAGCTCTGAAAGCAAAGTATCCAGACCTCTTTGAGGGTAAGGCCTAT	8280
2741	G Q Y L E A L K A K Y P D L F E G K A Y	2760
8281	GATTATTGGTATAAGAACTATGCAAATGATGGGTCAAATCCTTACTATACATTGTCACAC	8340
2761	D Y W Y K N Y A N D G S N P Y Y T L S H	2780
8341	GGTGACCGTGAATCTATCCCAGCAGATGTTGCTATTAAAGCAATGGTCAGCTAAGTATATG	8400
2781	G D R E S I P A D V A I K Q W S A K Y M	2800
8401	AACGGCAGAACGTTTTGGGCAATGGTATGGGTATGTATTGAAGGATPGGCATAATGGT	8460
2801	N G T N V L G N G M G Y V L K D W H N G	2820
8461	CAATATTTCAAGCTTGATGGTGATAAATCAACATTACCTCAAATTTAA	8508
2821	Q Y F K L D G D K S T L P Q I *	2835

25/37

Fig. 10

1	ATGAGAGACATGAGGGTAATTTGTGACCGTAAAAAATTGTACAAATCGGGCAAAGTACTA	60
1	M R D M R V I C D R K K L Y K S G K V L	20
61	GTAACAGCCGGTATTTTGTCTTTGATGATGTTTGGCGTCACAACTGCTAGTGTTAGTGCA	120
21	V T A G I F A L M M F G V T T A S V S A	40
121	AATACGATTGCAGTTGACACGAATCATAGCCGTACTTCAGCACAGATTAATAAGAGTGCC	180
41	N T I A V D T N H S R T S A Q I N K S A	60
181	GTTGATAAGGTAAATGATGACAAGACTACTTTAGGAGCGGCAAAAGTAGTGGCAGTAGCC	240
61	V D K V N D D K T T L G A A K V V A V A	80
241	ACAACGCCAGCGACACCGGTAGCAGATAAAACAGTAAGTGCACCCGCGCAGCAGATAAGGCA	300
81	T T P A T P V A D K T V S A P A A D K A	100
301	GTAGATACAACGTCATCAACGACACCTGCAACGGATAAGGCAGTAGATACAACGCCAACG	360
101	V D T T S S T T P A T D K A V D T T P T	120
361	ACACCTGCAGCAGATAAGGCAGTAGATACAACGCCAACGACACCTGCAGCAGATAAGGCA	420
121	T P A A D K A V D T T P T P A A D K A	140
421	GTAGATACAACGCCAACGACACCTGCAGCAAATAAAGCAGTAGATACAACGCCAGCGACC	480
141	V D T T P T T P A A N K A V D T T P A T	160
481	GCTGCAACAGATAAGGCGGTAGCCACGCCAGCCACACCTGCAGCAGATAAGCTAGCAAAT	540
161	A A T D K A V A T P A T P A A D K L A N	180
541	ACGACGCCCTGCAACGGACAAGGCAGTAGCCACAACGCCAGCGACGCCGGTAGCAAATAAA	600
181	T T P A T D K A V A T T P A T P V A N K	200
601	GCAGCAGACACGAGTAGTATTTCATGATCAACCATTAGATACAAATGTGCCAACTGATAAA	660
201	A A D T S S I H D Q P L D T N V P T D K	220
661	TCAGCAAACCTCGTCTCGACAACACAAAAAGTACGGATAATCAACAAGTTAAGTCTACA	720
221	S A N L V S T T Q K S T D N Q Q V K S T	240
721	GAAACATCTCATCTTCAAGAAATCAACGGTAAACCTATTTCTTGACGACAATGGTCAA	780
241	E T S H L Q E I N G K T Y F L D D N G Q	260
781	GTTAAAAAGAACTTCACCGCTATTATTGACGGTAAAGTTCTATACTTTGATAAAACATCC	840
261	V K K N F T A I I D G K V L Y F D K T S	280
841	GGCGAATTGACCGCAAATGCACCGCAAGTTACTAAGGGATTAGTAATATTGATAATGCA	900
281	G E L T A N A P Q V T K G L V N I D N A	300
901	CATAACGGCGCTCATGATCTCACAGCTGATAACTTCACAAATGTCGATGGTTACTTAACA	960
301	H N A A H D L T A D N F T N V D G Y L T	320
961	GCTAACAGTTGGTATCGTCCTAAGGACATCTTAAAAACGGAACGACCTGGACACCAACA	1020
321	A N S W Y R P K D I L K N G T T W T P T	340

1021	ACAGCAGAAGATTTTCGACCATGCTCATGTCTTGGTGGCCGGATAAGAATACGCAGGTA	1080
341	T A E D F R P L L M S W W P D K N T Q V	360
1081	GCTTATCTACAATATATGCAATCAGTTGGTATGCTACCTGACGATGTTAAAGTATCAAAT	1140
361	A Y L Q Y M Q S V G M L P D D V K V S N	380
1141	GATGATAATATGAGCACATTGACTGATGCTGCTATGACTGTTCAAAGAATATCGAATCG	1200
381	D D N M S T L T D A A M T V Q K N I E S	400
1201	CGAATGGTGTATCTGGAAAACTGATTGGCTCAAGCAAGATATGAACAACTGATTGAT	1260
401	R I G V S G K T D W L K Q D M N K L I D	420
1261	TCACAGGCAAAATGGAATATTGATAGTGAATCAAAGGGTAATGATCATTTACAGGGTGGG	1320
421	S Q A N W N I D S E S K G N D H L Q G G	440
1321	GCATTGTTATATGTGAATGATGACAAACACCTAACGCGAAGTACGATTACCGTCTGTTA	1380
441	A L L Y V N D D K T P N A N S D Y R L L	460
1381	AACCGTACACCAACCAACCAACCGGCCAAATTACTGATCCAAGTAAACAAGGTGGATAT	1440
461	N R T P T N Q T G Q I T D P S K Q G G Y	480
1441	GAGATGTTATTAGCTAATGATGTTGATAATTCTAACCTGTTGTACAAGCTGAGCAATTG	1500
481	E M L L A N D V D N S N P V V Q A E Q L	500
1501	AACGGCTTCACTACATGATGAACATTGGTACTATAGCTCAGAACGACCCACAGCTAAT	1560
501	N W L H Y M M N I G T I A Q N D P T A N	520
1561	TTTGACGGTTATCGTGTGATGCGGTTGATAACGTTGATGCCGATCTCTTACAAATTGCT	1620
521	F D G Y R V D A V D N V D A D L L Q I A	540
1621	GGTGATTACTTTAAAGCTGCATACGGTACTGGTAAAAGTGGGCAACGCAACCAATCAT	1680
541	G D Y F K A A Y G T G K T E A N A N N H	560
1681	ATTTGATCTTGAAGATTGGGATAATAATGATTCTGCGTACATTAAAGCCACGGGAAT	1740
561	I S I L E D W D N N D S A Y I K A H G N	580
1741	AACCAATTGACAATGGATTTTCCAGCACACTTGGCTTTGAAATACGCCTTGAACATGCCT	1800
581	N Q L T M D F P A H L A L K Y A L N M P	600
1801	CTTGCCGCACAAAGTGGCCTAGAACCGCTAATTAATACAAGTCTTGTTAAGCGTGGGAAA	1860
601	L A A Q S G L E P L I N T S L V K R G K	620
1861	GATGCCACAGAAAATGAAGCACAACTATGCCTTTATCCGTGCCATGATAGTGAA	1920
621	D A T E N E A Q P N Y A F I R A H D S E	640
1921	GTGCAGACCGTTATTGCACAAATTATTAAGGATAAAATTAACACAAATCAGACGGCTTA	1980
641	V Q T V I A Q I I K D K I N T K S D G L	660
1981	ACTGTAACACCAGATGAGATTAGCAAGCTTTCCTACTATTACAACGCCGATGAATTAAAA	2040
661	T V T P D E I K Q A F T I Y N A D E L K	680
2041	GCAGATAAGGAATATACAGCATACAATATTCCTGCTTCTTACGCTGTATTGTTGACAAAC	2100
681	A D K E Y T A Y N I P A S Y A V L L T N	700
2101	AAGGATACTGTGCCACGTGTTTATTATGTTGATCTATTTCTGATGATGGACAGTATATG	2160
701	K D T V P R V Y Y G D L F S D D G Q Y M	720
2161	TCACAGAAGTCACCATACTATGACGCCATTACGTCACCTTTTGAAGCCGTATCAAATAT	2220
721	S Q K S P Y Y D A I T S L L K S R I K Y	740

2221	GTTGCTGGTGGTCAAAGTATGAATATGACGTACTTGCATGAGTGCTTTGATCCAGCAAAA	2280
741	V A G G Q S M N M T Y L H E C F D P A K	760
2281	AATGAGACAAAGCCACAAGGTGTCTTAACATCAGTACGTTACGGTAAAGGTGCGATGACG	2340
761	N E T K P Q G V L T S V R Y G K G A M T	780
2341	GCTGACGATTTGGGTAATAGTGACACACGTCAACAAGGTATTGGTTTGGTGATTAATAAT	2400
781	A D D L G N S D T R Q Q G I G L V I N N	800
2401	AAGCCATTCTTGAATTTAAATGATGATGAACAAATGTGCTCAATATGGGTGCTGCTCAC	2460
801	K P F L N L N D D E Q I V L N M G A A H	820
2461	AAAAATCAAGCTTACCGACCACTTATGTTGACAAACAAATCTGGTCTTCAAATTTACGAT	2520
821	K N Q A Y R P L M L T T K S G L Q I Y D	840
2521	AAGGATGCCGGAGCGCCAGTTGTTTATACTAACGATGCTGGTCAACTTATTTTTAAGTCA	2580
841	K D A G A P V V Y T N D A G Q L I F K S	860
2581	GATATGGTCTATGGTGTGACGAATCCACAGGTATCTGGTTATTTTGCTGCATGGGTACCA	2640
861	D M V Y G V S N P Q V S G Y F A A W V P	880
2641	GTCGGTGCGAGTGATAGTCAAGATGCTAGAACACAAAGCAGCCAGTCAGAACTAAGGAT	2700
881	V G A S D S Q D A R T Q S S Q S E T K D	900
2701	GGCGATGTCTATCATTCAAATGCTGCGCTTGATTCTAATGTGATTTATGAAGGCTTCTCG	2760
901	G D V Y H S N A A L D S N V I Y E G F S	920
2761	AATTTCCAAGCAATGCCTGAAAAGATGATGACTTCACCAACGTAAAAATTGCTCAAAAT	2820
921	N F Q A M P E K N D D F T N V K I A Q N	940
2821	GCTAAATTGTTTAAAGATTTAGGATTACAAGCTTTGAATTAGCACCGCAATATCGTTCA	2880
941	A K L F K D L G I T S F E L A P Q Y R S	960
2881	AGTACAGATAATAGTTTTTTGGATTGCGTTATCCAAAACGGCTATGCCCTTTACTGATCGA	2940
961	S T D N S F L D S V I Q N G Y A F T D R	980
2941	TATGATGTTGGCTATAATACGCCAACAAATATGGTACAGTTGATCAACTTCTAGATAGT	3000
981	Y D V G Y N T P T K Y G T V D Q L L D S	1000
3001	CTAAGAGCATTACACGCACAAGGTATTCAGGCTATTAATGACTGGGTACCTGATCAAATT	3060
1001	L R A L H A Q G I Q A I N D W V P D Q I	1020
3061	TATAATTTACCTGGCGAACAAATCGTCACCGCAGTTCGTACAAATGGTTCAGGTAAGTAC	3120
1021	Y N L P G E Q I V T A V R T N G S G K Y	1040
3121	GATTATGATTCAAGTGATTAATAACACGCTCTATGATTACGAACAGTTGGGGGCGGCGAA	3180
1041	D Y D S V I N N T L Y D S R T V G G G E	1060
3181	TACCAAGAAAAGTTTGGTGGCCTGTTCTTAGACAGTTGAAAAAGATTATCCTAGCTTG	3240
1061	Y Q E K F G G L F L D Q L K K D Y P S L	1080
3241	TTTGAACTAAGCAGATATCAACGAATCAGCCGATGAATCCGGATGTTAAAATTAAAGAA	3300
1081	F E T K Q I S T N Q P M N P D V K I K E	1100
3301	TGGTCTGCAAAGTACTTTAATGGTTCAAACATTCAAGGTCGTGGCGCTTGGTATGTACTT	3360
1101	W S A K Y F N G S N I Q G R G A W Y V L	1120
3361	AAAGACTGGGCAACAAATCAATATTTCAATGTGTCTAGTGATAATGGATTCTTGCCTAAA	3420
1121	K D W A T N Q Y F N V S S D N G F L P K	1140

3421	CAGTTACTGGGTGAAAAACAAGCACCGGCTTTATAACAGAAATGGTAAGACTTCTTTC	3480
1141	Q L L G E K T S T G F I T E N G K T S F	1160
3481	TACTCAACAAGTGGTTATCAAGCTAAAGATACCTTTATTCAAGATGGAACAAATTGGTAT	3540
1161	Y S T S G Y Q A K D T F I Q D G T N W Y	1180
3541	TACTTTGATAATGCAGGCTATATGTTGACAGGTAAACAAAATATCCACGATAAAAATTAT	3600
1181	Y F D N A G Y M L T G K Q N I H D K N Y	1200
3601	TATTTCTTACCTAATGGTGTGGAACCTCAAGATGCTTACCTTTTGTATGGTAATCAAGAA	3660
1201	Y F L P N G V E L Q D A Y L F D G N Q E	1220
3661	TTTACTATAATAAGCTGGGGAACAAGTTATGAACCAGTATTATCAAGATAGTCAAAT	3720
1221	F Y Y N K A G E Q V M N Q Y Y Q D S Q N	1240
3721	CAATGGCATTATTTCTTTGAAAATGGTCGCATGGCAATTGGCCTGACAGAAGTTCCGAAC	3780
1241	Q W H Y F F E N G R M A I G L T E V P N	1260
3781	GCTGATGGCACCCATGTTACACAATATTTTGTATGCTAATGGTGTCCAAATTAAGGCACA	3840
1261	A D G T H V T Q Y F D A N G V Q I K G T	1280
3841	GCTATAAAAGATCAGAATAATCAATTACGCTATTTTGTATGAGGCCACAGGTAATATGGTG	3900
1281	A I K D Q N N Q L R Y F D E A T G N M V	1300
3901	GTTAATTCATGGGGACAGTTAGCAGATAAGTCTTGGCTTTACCTTAATGCACAAGGCGTT	3960
1301	V N S W G Q L A D K S W L Y L N A Q G V	1320
3961	GCTGTGACTGGTAACCAAAAAATTGATGGTGAAGAGTACTACTTCAATGCTGATGGTAAG	4020
1321	A V T G N Q K I D G E E Y Y F N A D G K	1340
4021	CAAGTTAAAGGCAATGCAATCATCGATAATAATGGTGATCAACGTTATTATGATGGTGAT	4080
1341	Q V K G N A I I D N N G D Q R Y Y D G D	1360
4081	AAGGGTGTATGGTAGTTAATTCATGGGGTGAGTTGCCAGATGGCTCATGGTTATATTG	4140
1361	K G V M V V N S W G E L P D G S W L Y L	1380
4141	AATGACAAAGGTATTGCTGTAACAGGCCGTCAAGTCATTAATAATCAAGTTAATTTCTTT	4200
1381	N D K G I A V T G R Q V I N N Q V N F F	1400
4201	GGTAATGATGGTAAGCAAATCAAAGATGCCTTTAAATTATTATCCGATGGTTCATGGGTG	4260
1401	G N D G K Q I K D A F K L L S D G S W V	1420
4261	TATTTGGATGATAAGGGCCTGATAACAACCTGGAGCCAAAGTTATCAATGGTCTAAATATG	4320
1421	Y L D D K G L I T T G A K V I N G L N M	1440
4321	TTTTTTGATAAAGACGGTCATCAAATCAAAGGTGATGCCAGCACGGATGCCAATGGTAAG	4380
1441	F F D K D G H Q I K G D A S T D A N G K	1460
4381	CGCCATTATTATGACAAAAATGATGGTCATCTTGTCACAAATTCATGGGGTGAGTTGCCA	4440
1461	R H Y Y D K N D G H L V T N S W G E L P	1480
4441	GATGGTTCATGGTTATATCTAGAAGAACAAGGTGATGCTGTTACTGGTCAACGTGTGATT	4500
1481	D G S W L Y L E E Q G D A V T G Q R V I	1500
4501	GATGGCAAGACACGCTATTTTGTATGAAGATGGCAAACAAATTAATAAGCCTAAAAACG	4560
1501	D G K T R Y F D E D G K Q I K N S L K T	1520
4561	CTGGCCAATGGCGATAAGATTTATCTTGATGGTGATGGGGTTGCTGCAACAGGCTTACAA	4620
1521	L A N G D K I Y L D G D G V A A T G L Q	1540



4621	CATGTGGGCGATAAAATCATGTATTTTGATGAAGATGGCAAACAAGTTGTTGGCAAGTTT	4680
1541	H V G D K I M Y F D E D G K Q V V G K F	1560
4681	GATCAGCAAAGATGGTTCATGGTATTACTTAAATCAGGATGGTGTGCCGCGGTTGGT	4740
1561	V S A K D G S W Y Y L N Q D G V A A V G	1580
4741	CCAAGCAGCATTAAATGGACAATCACTTTACTTTGATCAAGATGGTAAACAAGTTAAATAT	4800
1581	P S S I N G Q S L Y F D Q D G K Q V K Y	1600
4801	AATGAAGTTCGTAATAGTGATGGAACAACCAACTATTACACAGGATTAACGGGTGAAAAG	4860
1601	N E V R N S D G T T N Y Y T G L T G E K	1620
4861	TTAACGCAAGACTTCGGTGAACCTACCAGATGGTTCATGGATTATCTTGATGCGCAAGGT	4920
1621	L T Q D F G E L P D G S W I Y L D A Q G	1640
4921	CATACAGTAACTGGTGACAAATCATTAAACGGTCAAAATCTTTACTTTAAGGCTGACGGC	4980
1641	H T V T G A Q I I N G Q N L Y F K A D G	1660
4981	CAGCAAGTTAAAGGTCATGCTTATACTGACCAATTAGGTCATATGCGTTTTATGATCCT	5040
1661	Q Q V K G H A Y T D Q L G H M R F Y D P	1680
5041	GATTCAGGTGATATGTTGAGTAATCGCTTTGAACAAATCACACCTGGTGTATGGGCTTAC	5100
1681	D S G D M L S N R F E Q I T P G V W A Y	1700
5101	TTTGGTGCTGATGGTGTGGCCATAACTGGACAACATGACATAAATGGTCAGAAGCTATTC	5160
1701	F G A D G V A I T G Q H D I N G Q K L F	1720
5161	TTTGATGAGACAGGATATCAAGTTAAAGGTTTCGCAACGTACAATAGATGGTACGTTATAC	5220
1721	F D E T G Y Q V K G S Q R T I D G T L Y	1740
5221	AGCTTCGATTCTCAAACCTGGTAACCAAAAACGCGTACAGACACATTTGTTGCCACAAGCA	5280
1741	S F D S Q T G N Q K R V Q T T L L P Q A	1760
5281	GGTCACTATATCACGAAAAATGGTAACGATTGGCAGTATGATACCAATGGTGAACCTAGCG	5340
1761	G H Y I T K N G N D W Q Y D T N G E L A	1780
5341	AAGGGTCTGCGTCAAGATAGCAATGGTAAGTTGCGTTACTTTGATTTGACAACCGGCATA	5400
1781	K G L R Q D S N G K L R Y F D L T T G I	1800
5401	CAAGCGAAAGGCCAATTTGTTACAATTGGCCAAGAACTTATTACTTTAGTAAAGATCAC	5460
1801	Q A K G Q F V T I G Q E T Y Y F S K D H	1820
5461	GGGGATGCGCAGTTATTGCCAATGGTCACTGAAGGGCATTACGGTACAATAACACTCAAG	5520
1821	G D A Q L L P M V T E G H Y G T I T L K	1840
5521	CAAGGTCAAGACACCAAAAACAGCCTGGGTTTACCGTGATCAAAATAATACTATTTTGAAG	5580
1841	Q G Q D T K T A W V Y R D Q N N T I L K	1860
5581	GGATTGCAAAATATCAATGGCACGTTGCAATTCTTTGATCCATATACAGGTGAACAACTT	5640
1861	G L Q N I N G T L Q F F D P Y T G E Q L	1880
5641	AAGGGTGGCGTAGCAAAGTATGACGACAAGCTCTTTACTTTGAATCAGGTAAAGGTAAT	5700
1881	K G G V A K Y D D K L F Y F E S G K G N	1900
5701	CTTGTTAGCACCGTAGCAGGTGACTATCAGGATGGTCATTATATTCCCAAGATGGCCAA	5760
1901	L V S T V A G D Y Q D G H Y I S Q D G Q	1920
5761	ACACGTTACGCAGATAAGCAAAATCAGCTTGTAAGGGGACTTGTTACTGTTAATGGGGCA	5820
1921	T R Y A D K Q N Q L V K G L V T V N G A	1940

5821	TTACAATACTTTGATAACGCTACTGGTAACCAATAAAAAATCAACAAGTTATTGTTGAT	5880
1941	L Q Y F D N A T G N Q I K N Q Q V I V D	1960
5881	GGCAAGACGTACTATTTTGACGATAAAGGCAATGGTGAATACTTATTCATAATACATTA	5940
1961	G K T Y Y F D D K G N G E Y L F T N T L	1980
5941	GATATGTCTACTAATGCTTTTTTCTACCAAAATGTTGCATTCAATCATGACAGTAGCAGT	6000
1981	D M S T N A F S T K N V A F N H D S S S	2000
6001	TTCGACCATACTGTTGATGGCTTCTTGACGGCAGATACTTGGTATCGACCAAAGTCAATT	6060
2001	F D H T V D G F L T A D T W Y R P K S I	2020
6061	TTGGCTAACGGGACAACCTTGGCGTGATTGCGACTGATAAGGATATGCGACCATTAACTACT	6120
2021	L A N G T T W R D S T D K D M R P L I T	2040
6121	GTTTGGTGGCCAAATAAGAATGTTCAAGTCAACTACCTCAACTTCATGAAAGCAAATGGC	6180
2041	V W W P N K N V Q V N Y L N F M K A N G	2060
6181	TTGTTGACAACAGCAGCACAAATACACACTACATTGAGATCAATATGATTGAACCAAGCT	6240
2061	L L T T A A Q Y T L H S D Q Y D L N Q A	2080
6241	GCACAAGATGTTCAAGTGGCCATTGAAAGCGCATTCGTCAGAGCATGGCACAGACTGG	6300
2081	A Q D V Q V A I E R R I A S E H G T D W	2100
6301	TTACAGAAATGTTGTTTGAATCACAAAATAAATACCCATCATTGTTGAAGCAACAATTC	6360
2101	L Q K L L F E S Q N N N P S F V K Q Q F	2120
6361	ATTTGGAACAAGGATTCTGAATATCATGGTGGTGGTATGCTTGGTTCCAAGTGGTTAT	6420
2121	I W N K D S E Y H G G G D A W F Q G G Y	2140
6421	CTGAAGTATGGCAATAACCCACTCACACCAACAACTAATTCTGATTATCGTCAACCTGGT	6480
2141	L K Y G N N P L T P T T N S D Y R Q P G	2160
6481	AATGCATTGATTCTTCTGCTAGCCAACGACGTGGATAATTCTAATCCTGTTGTGCAAGCT	6540
2161	N A F D F L L A N D V D N S N P V V Q A	2180
6541	GAAAACCTAAACTGGTTACATTACTTAATGAACTTTGGCACCATCACTGCGGGTCAAGAT	6600
2181	E N L N W L H Y L M N F G T I T A G Q D	2200
6601	GACGCTAATTTTGATAGTATTCGTATTGACGCTGTGCGACTTTATTCATAATGATACAATC	6660
2201	D A N F D S I R I D A V D F I H N D T I	2220
6661	CAACGTACTTATGATTATCTTCGTGATGCTTATCAAGTGCAACAAAGTGAAGCCAAAGCA	6720
2221	Q R T Y D Y L R D A Y Q V Q Q S E A K A	2240
6721	AACCAGCACATTTTATTGGTTGAAGCTGGCTTAGACGCAGGTACATCAACGATTGATAAT	6780
2241	N Q H I S L V E A G L D A G T S T I H N	2260
6781	GATGCGTTAATTGAGTCAAACTCCGTGAAGCAGCGACATTGTCGTTAACAATGAACCT	6840
2261	D A L I E S N L R E A A T L S L T N E P	2280
6841	GGTAAAAATAAACCATTTGACGAATATGCTACAAGACGTTGACGGCGGTACGCTTATCACC	6900
2281	G K N K P L T N M L Q D V D G G T L I T	2300
6901	GACCATACGCAGAATAGTACAGAAAATCAGGCGACACCAAATATTCAATTATTCACGCG	6960
2301	D H T Q N S T E N Q A T P N Y S I I H A	2320
6961	CACGATAAAGGTGTGCAAGAAAAAGTAGGTGCAGCCATTACTGATGCTACTGGTGCTGAT	7020
2321	H D K G V Q E K V G A A I T D A T G A D	2340

7021	TGGACGAACTTTACAGATGAACAGTTAAAGCCGGATTAGAGCTATTCTATAAGGATCAG	7080
2341	W T N F T D E Q L K A G L E L F Y K D Q	2360
7081	CGCGCAACAAACAAAAGTATAATAGTTATAACATACCAAGTATTTATGCCCTGATGTTG	7140
2361	R A T N K K Y N S Y N I P S I Y A L M L	2380
7141	ACAAACAAAGATACTGTTCTCGTATGTATTATGGGGATATGTATCAAGATGACGGACAG	7200
2381	T N K D T V P R M Y Y G D M Y Q D D G Q	2400
7201	TATATGGCAAACAAGAGTATCTACTATGATGCCTTAGTGTCTTAATGACGGCTCGTAAA	7260
2401	Y M A N K S I Y Y D A L V S L M T A R K	2420
7261	AGCTATGTGACGGTGGTCAAACCTATGAGTGTGACAATCATGGTTTGTGAAGAGTGTG	7320
2421	S Y V S G G Q T M S V D N H G L L K S V	2440
7321	CGTTTTCGAAAAGATGCGATGACAGCTAATGATTTAGGTACATCAGCTACGCGTACTGAG	7380
2441	R F G K D A M T A N D L G T S A T R T E	2460
7381	GGTCTTGGTGTCTATTATTGGTAATGATCCAAAGTTGCAACTTAATGATTCGGATAAAGTG	7440
2461	G L G V I I G N D P K L Q L N D S D K V	2480
7441	ACACTGGATATGGGTGACGACATAAAAATCAAAGTATCGCGCAGTTATCTTAACAACA	7500
2481	T L D M G A A H K N Q K Y R A V I L T T	2500
7501	CGTGATGGTTTGGCAACCTTTAATTCAGATCAAGCACCAACAGCTTGGACAAACGATCAA	7560
2501	R D G L A T F N S D Q A P T A W T N D Q	2520
7561	GGAACGTTAACATTCTCAAATCAAGAGATTAACGGGCAAGACAATACACAAATTCGTGGT	7620
2521	G T L T F S N Q E I N G Q D N T Q I R G	2540
7621	GTTGCTAATCCGCAAGTTTCTGGTTATCTAGCTGTTTGGGTGCCTGTGGGTGCATCAGAC	7680
2541	V A N P Q V S G Y L A V W V P V G A S D	2560
7681	AATCAAGATGCCCCGTACAGCAGCAACGACAACAGAAAATCATGATGGTAAAGTATTACAC	7740
2561	N Q D A R T A A T T T E N H D G K V L H	2580
7741	TCGAATGCGGCATTAGATTCTAACCTTATTTATGAAGGTTTCTCTAACTTCCAACCTAAG	7800
2581	S N A A L D S N L I Y E G F S N F Q P K	2600
7801	GCAACAACGCATGATGAACCTTACGAACGTTGTAATTGCTAAAAATGCCGATGTCTTCAAT	7860
2601	A T T H D E L T N V V I A K N A D V F N	2620
7861	AATTGGGGTATTACAGTTTTTGAATGGCACCACAGTACCGTTCAAGTGGGGACCATACA	7920
2621	N W G I T S F E M A P Q Y R S S G D H T	2640
7921	TTCTTGGATTCAACGATTGATAATGGTTATGCCTTCACTGATCGCTATGACTTAGGTTTC	7980
2641	F L D S T I D N G Y A F T D R Y D L G F	2660
7981	AATACACCAACAAAGTATGGCACTGATGGTGATTGCGTGCAACGATTCAAGCGCTACAT	8040
2661	N T P T K Y G T D G D L R A T I Q A L H	2680
8041	CATGCTAATATGCAAGTTATGGCTGACGTTGTTGATAACCAGGTCTATAACTTACCTGGT	8100
2681	H A N M Q V M A D V V D N Q V Y N L P G	2700
8101	AAAGAAGTTGTTTCAGCAACACGACGAGGTGTTTATGGTAATGACGACGCCACGGGCTTT	8160
2701	K E V V S A T R A G V Y G N D D A T G F	2720
8161	GGAACGCAACTCTATGTGACTAACTCCGTTGGTGGTGGTCAATACCAAGAGAAATATGCT	8220
2721	G T Q L Y V T N S V G G G Q Y Q E K Y A	2740

8221	GGACAATACTTAGAAGCTCTGAAAGCAAAGTATCCAGACCTCTTTGAGGGTAAGGCCTAT	8280
2741	G Q Y L E A L K A K Y P D L F E G K A Y	2760
8281	GATTATTGGTATAAGAACTATGCAAATGATGGGTCAAATCCTTACTATACATTGTCACAC	8340
2761	D Y W Y K N Y A N D G S N P Y Y T L S H	2780
8341	GGTGACCGTGAATCTATCCCAGCAGATGTTGCTATTAAGCAATGGTCAGCTAAGTATATG	8400
2781	G D R E S I P A D V A I K Q W S A K Y M	2800
8401	AACGGCACGAACGTTTTGGGCAATGGTATGGTTATGTATTGAAGGATTGGCATAATGGT	8460
2801	N G T N V L G N G M G Y V L K D W H N G	2820
8461	CAATATTTCAAGCTTGATGGTGATAAATCAACATTACCTCAAATTT	8506
2821	Q Y F K L D G D K S T L P Q I *	2835

Fig. 11

Séquence du gène *dsrD*

1 AATAATCTGT CTCCATTGCT TTCAAAATAA TAATAGTTAA TTATTATCAT  
 51 GGAACAATCA ATATTTTATT TATATTCACT ATTGAATATC CTTTTTTGCA  
 101 TAAATCTCTA GAGCCGATT TTTGGGTTAT ACAATGAATT GGTAAAGGTT  
 151 AATCATTTTT ACAAACCAT GGTGGTTTTT TATTTTTTCT AAAATTACCG  
 201 AACTAGAGGA AGAGAAAAGG AGCAATAGTT GTATGAGAGA CATGAGGGTA  
 251 ATTTGTGACC GTAAAAAATT GTACAAATCG GGCNAAGTAC TAGTAACAGC  
 301 CGGTATTTTT GCTTTGATGA TGTTTGGCGT CACAACGCT AGTGTTAGTG  
 351 CAAATACGAT TGCAGTTGAC ACGAATCATA GCCCTACTTC AGCACAGATT  
 401 AATAAGAGTG CCGTTGATAA GGTAAATGAT GACAAGACTA CTTTAGGAGC  
 451 GGCAAAAGTA GTGGCAGTAG CCACAACGCC AGCGACACCG GTAGCAGATA  
 501 AAACAGTAAG TGCACCCGCA GCAGATAAGG CAGTAGATAC AACGTCAATCA  
 551 ACGACACCTG CAACGGATAA GGCAGTAGAT ACAACGCCAA CGACACCTGC  
 601 AGCAGATAAG GCAGTAGATA CAACGCCAAC GACACCTGCA GCAGATAAGG  
 651 CAGTAGATAC AACGCCAAGC ACACCTGCAG CAAATAAGC AGTAGATACA  
 701 ACGCCAGCEA CCGCTGCAAC AGATAAGGCG GTAGCCACGC CAGCCACACC  
 751 TGCAGCAGAT AAGCTAGCAA ATACGACGCC TGCAACGGAC AAGGCAGTAG  
 801 CCACAACGCC AGCGACGCCG GTAGCAANTA AAGCAGCAGA CAGGAGTAGT  
 851 ATTCAATGATC AACCATTAGA TACAAATGTG CCAACTGATA AATCAGCAAA  
 901 CCTCGTCTCG AGAACACAAA AAGTACGGA TAATCAACAA GTTAAGTCTA  
 951 CAGAAACATC TCATCTTCAA GAAATCAACG GTAAAACCTA TTTTCTTGAC  
 1001 GACAATGGTC AAGTTAAAAA GAACTTCACC GCTATTATTG ACGGTAAAGT  
 1051 TCTATACTTT GATAAAACAT CCGGCGAATT GACCGCAAT GCACCGCAAG  
 1101 TTACTAAGGG ATTAGTAAAT ATTGATAATG CACATAACGC GGCTCATGAT  
 1151 CTCACAGCTG ATAACTTCAC AAATGTCEAT GGTACTTAA CAGCTAACAG  
 1201 TTGGTATCGT CCTAAGGACA TCTTAAAAAA CGGAACGACC TGGACACCAA  
 1251 CAACAGCAGA AGATTTTCGA CCATTGCTCA TGTCTTGGTG GCCGGATAAG  
 1301 AATACGCAGG TAGCTTATCT ACAATATATG CAATCAGTTG GTATGCTACC  
 1351 TGACGATGTT AAGTATCAA ATGATGATAA TATGAGCACA TTGACTGATG  
 1401 CTGCTATGAC TGTCAAAAG AATATCGAAT CGCGAATTGG TGTATCTGGA

3001 AGCAATGCCT GAAAAGAATG ATGACTTCAC CAACGTAAAA ATTGCTCAAA  
 3051 ATGCTAAATT GTTTAAAGAT TTAGGGATTA CAAGCTTTGA ATTAGCACCG  
 3101 CAATATCGTT CAAGTACAGA TAATAGTTTT TTGGATTCCG TTATCCAAA  
 3151 CGGCTATGCC TTTACTGATC GATATGATGT TGGCTATAAT ACGCCAACAA  
 3201 AATATGGTAC AGTTGATCAA CTTCTAGATA GTCTAAGAGC ATTACACGCA  
 3251 CAAGGTATTC AGGCTATTAA TGA CTGGGTA CCTGATCAAA TTTATAATTT  
 3301 ACCTGGCGAA CAAATCGTCA CCGCAGTTCC TACAAATGGT TCAGGTAAGT  
 3351 ACGATTAATGA TTCAGTGAAT AATAACACGC TCTATGATTC ACGAACAGTT  
 3401 GGGGGCGGCG AATACCAAGA AAAGTTTGGT GGCTGTCTCT TAGACCAGTT  
 3451 GAAAAAGAT TATCCTAGCT TGT TTGAAC TAAGCAGATA TCAACGAATC  
 3501 AGCCGATGAA TCCGGATGTT AAAATTAAAG AATGGTCTGC AAAGTACTTT  
 3551 AATGGTTCAA ACATTCAAGG TCGTGGCGCT TGGTATGTAC TTAAGACTG  
 3601 GGCAACAAAT CAATATTTCA ATGTGTCTAG TGATAATGGA TTCTTGCCTA  
 3651 AACAGTTACT GGTGAAAAA ACAAGCACCG GCTTTATAAC AGAAAATGGT  
 3701 AAGACTTCTT TCTACTCAAC AAGTGGTTAT CAAGCTAAG ATACCTTTAT  
 3751 TCAAGATGGA ACAAATTGGT ATTACTTTGA TAATGCAGGC TATATGTTGA  
 3801 CAGGTAAACA AAATATCCAC GATAAAAAAT ATTATTTCTT ACCTAATGGT  
 3851 GTGGAACCTC AAGATGCTTA CCTTTTGTAT GGTAAATCAAG AATTTTACTA  
 3901 TAATAAGCT GGGGAACAAG TTATGAACCA GTATTATCAA GATAGTCAAA  
 3951 ATCAATGGCA TTATTTCTTT GAAAATGGTC GCATGGCAAT TGGCCTGACA  
 4001 GAAGTTCCGA ACGCTGATGG CACCCATGTT ACACAATATT TTGATGCTAA  
 4051 TGGTGTCCAA ATTAAAGGCA CAGCTATAAA AGATCAGAAAT AATCAATTAC  
 4101 GCTATTTTGA TGAGGCCACA GGTAAATATG TGGTTAATTC ATGGGGACAG  
 4151 TTAGCAGATA AGTCTTGGCT TTACCTTAAT GCACAAGGCG TTGCTGTGAC  
 4201 TGGTAACCAA AAAATTGATG GTGAAGAGTA CTACTTCAAT GCTGATGGTA  
 4251 AGCAAGTTAA AGGCAATGCA ATCATCGATA ATAATGGTGA TCAACGTTAT  
 4301 TATGATGGTG ATAAGGGTGT CATGGTAGTT AATTCATGGG GTGAGTTGCC  
 4351 AGATGGCTCA TGGTTATATT TGAATGACAA AGGATTGCT GTAACAGGCC  
 4401 GTCAAGTCAT TAATAATCAA GTTAATTTCT TTGGTAATGA TGGTAAGCAA  
 4451 ATCAAGATG CCTTAAATTT ATATCCGAT GGTCATGGG TGTATTTGGA

4501 TGATAAGGGC CTGATAACAA CTGGAGCCAA AGTTATCAAT GGTCTAAATA  
 4551 TGTTTTTTGA TAAAGACGGT CATCAAATCA AAGGTGATGC CAGCACGGAT  
 4601 GCCAATGGTA AGCGCCATTA TTATGACAAA AATGATGGTC ATCTTGTCAC  
 4651 AAATTCATGG GGTGAGTTGC CAGATGGTTC ATGCTTATAT CTAGAAGAAC  
 4701 AAGGTGATGC TGTACTGGT CAACGTGTGA TTGATGGCAA GACACGCTAT  
 4751 TTGATGAAG ATGGCAAACA AATTAAAAAT AGCCTAAAAA CGCTGGCCAA  
 4801 TGGCGATAAG ATTTATCTTG ATGGTGATGG GGTGCTGCA ACAGGCTTAC  
 4851 AACATGTGGG CGATAAAATC ATGTATTTTG ATGAAGATGG CAAACAAGTT  
 4901 GTTGGCAAGT TTGTATCAGC AAAAGATGGT TCATGGTATT ACTTAAATCA  
 4951 GGATGGTGTT GCCGCGGTTG GTCCAAGCAG CATTAAATGA CAATCACTTT  
 5001 ACTTTGATCA AGATGGTAAA CAAGTTAAAT ATAATGAAGT TCGTAATAGT  
 5051 GATGGAACAA CCAACTATTA CACAGGATTA ACGGGTGAAA AGTTAACGCA  
 5101 AGACTTCGGT GAACTACCAG ATGGTTCATG GATTTATCTT GATGCGCAAG  
 5151 GTCATACAGT AACTGGTGCA CAAATCATT ACGGTCAAAA TCTTTACTTT  
 5201 AAGGCTGACG GCCAGCAAGT TAAAGGTCAT GCTTATACTG ACCAATAGG  
 5251 TCATATGCGT TTTTATGATC CTGATTCAGG TGATATGTTG AGTAATCGCT  
 5301 TTGAACAAAT CACACCTGGT GTATGGGCTT ACTTTGGTGC TGATGGTGTG  
 5351 GCCATAACTG GACAACATGA CATAAATGGT CAGAAGCTAT TCTTTGATGA  
 5401 GACAGGATAT CAAGTTAAAG GTTCGCAACG TACAATAGAT GGTACGTTAT  
 5451 ACAGCTTCGA TTCTCAAAC TGTAAACCAA AACGCGTACA GACAACATG  
 5501 TTGCCACAAG CAGGTCCTA TATCACGAAA AATGGTAACG ATTGGCAGTA  
 5551 TGATAACCAAT GGTGAAC TAGAAGGGTCT GCGTCAAGAT AGCAATGGTA  
 5601 AGTTGCGTTA CTTTGATTTG ACAACCGGCA TACAAGCGAA AGGCCAATTT  
 5651 GTTACAATTG GCCAAGAAAC TTATTACTTT AGTAAAGATC ACGGGCATGC  
 5701 GCAGTTATTG CCAATGGTCA CTGAAGGGCA TTACGGTACA ATAACACTCA  
 5751 AGCAAGGTCA AGACACCAAA ACAGCCTGGG TTACCGTGA TCAAAATAAT  
 5801 ACTATTTTGA AGGGATTGCA AAATATCAAT GGCACGTTGC AATTCTTTGA  
 5851 TCCATATACA GGTGAACAAC TTAAGGGTGG CGTAGCAAAG TATGACGACA  
 5901 AGCTCTTTTA CTTTGAATCA GGTAAAGGTA ATCTTGTTAG CACCGTAGCA  
 5951 GGTGACTATC AGGATGGTCA TTATATTTCC CAAGATGGCC AAACACGTTA  
 6001 CGCAGATAAG CAAATCAGC TTGTAAAGGG ACTTGTTCT GTTAATGGGG

6051 CATTACAATA CTTTGATAAC GCTACTGGTA ACCAATAAA AAATCAACAA  
 6101 GTTATTGTG ATGGCAAGAC GTACTATTTT GACGATAAAG GCAATGGTGA  
 6151 ATACTTATTC ACTAATACAT TAGATATGTC TACTAATGCT TTTTCTACCA  
 6201 AAAATGTTGC ATTCAATCAT GACAGTAGCA GTTTCGACCA TACTGTTGAT  
 6251 GGCTTCTTGA CGGCAGATAC TTGGTATCEA CCAAAGTCAA TTTTGGCTAA  
 6301 CGGGACAACCT TGGCGTGATT CGACTGATAA GGATATGCGA CCATTAATCA  
 6351 CTGTTTGGTG GCCAAATAAG AATGTTCAAG TCAACTACCT CAACTTCATG  
 6401 AAAGCAAATG GCTTGTTCAC AACAGCAGCA CAATACACAC TACATTCCAA  
 6451 TCAATATGAT TTGAACCAAG CTGCACAAGA TGTTCAAGTG GCCATTGAAA  
 6501 GGGCGATTGC GTCAGAGCAT GGCACAGACT GGTACAGAA ATTGTTGTTT  
 6551 GAATCACAAA ATAATAACCC ATCATTGTG AAGCAACAAT TCATTTGGAA  
 6601 CAAGGATTCT GAATATCATG GTGGTGGTGA TGCTTGGTTC CAAGGTGOTT  
 6651 ATCTGAAGTA TGGCAATAAC CCACTCACAC CAACAATAA TTCTGATTAT  
 6701 CGTCAACCTG GTAATGCATT TGATTTCCTG CTAGCCAACG ACGTGGATAA  
 6751 TTCTAATCCT GTTGTGCAAG CTGAAAACCTT AAAGTGGTTA CATTACTTAA  
 6801 TGAACTTTGG CACCATCACT GCGGGTCAAG ATGACGCTAA TTTTGATAGT  
 6851 ATTGCTATTG ACGCTGTCTGA CTTTATTCAT AATGATACAA TCCAACGTAC  
 6901 TTATGATTAT CTTGCTGATG CTTATCAAGT GCACCAAGT GAAGCCAAAG  
 6951 CAAACCAGCA CATTTCATG GTTGAAGCTG GCTTAGACGC AGGTACATCA  
 7001 ACGATTCAAT ATGATGCGTT AATTGAGTCA AACCTCCGTG AAGCAGCGAC  
 7051 ATTGTCGTTA ACAAATGAAC CTGGTAAAAA TAAACCATG ACGAATATGC  
 7101 TACAAGACGT TGACGGCGGT ACGCTTATCA CCGACCATAC GCAGAATAGT  
 7151 ACAGAAAATC AGGCGACACC AACTATTCA ATTATTACCG CGCAGCATAA  
 7201 AGGTGTGCAA GAAAAAGTAG GTGCAGCCAT TACTGATGCT ACTGGTGCTG  
 7251 ATTGGACGAA CTTTACAGAT GAACAGTTAA AAGCCGGATT AGAGCTATTC  
 7301 TATAAGGATC AGCGCGCAAC AAACAAAAAG TATAATAGTT ATAACATACC  
 7351 AAGTATTTAT GCCCTGATGT TGACAAACAA AGATACTGTT CCTCGTATGT  
 7401 ATTATGGGGA TATGTATCAA GATGACGGAC AGTATATGGC AAACAAGAGT  
 7451 ATCTACTATG ATGCCTTAGT GTCATTAATG ACGGCTCGTA AAAGCTATGT  
 7501 CAGCGGTGGT CAACTATGA GTGTTGACAA TCATGGTTG TTGAAGAGTG



7551 TCCGTTTTGG AAAAGATGCG ATGACAGCTA ATGATTTAGG TACATCAGCT  
 7601 ACGCGTACTG ACGGTCTTGG TGTCAATTAT GGTAATGATC CAAAGTTGCA  
 7651 ACTTAATGAT TCGGATAAAG TGACACTGGA TATGGGTGCA GCACATAAAA  
 7701 ATCAAAAAGTA TCGCGCAGTT ATCTTAACAA CACGTGATGG TTTGGCAACC  
 7751 TTTAATTCAG ATCAAGCACC AACAGCTTGG ACAACCGATC AAGGAACGTT  
 7801 AACATTCTCA AATCAAGAGA TTAACGGGCA AGACAATACA CAAATTCGTG  
 7851 GTGTTGCTAA TCCGCAAGTT TCTGGTTATC TAGCTGTTTG GGTGCCTGTG  
 7901 GGTGCATCAG ACAATCAAGA TGCCCGTACA GCAGCAACGA CAACAGAAAA  
 7951 TCATGATGCT AAAGTATTAC ACTCGAATGC GGCATTAGAT TCTAACCTTA  
 8001 TTTATGAAGG TTTCTCTAAC TTCCAACCTA AGGCAACAAC GCATGATGAA  
 8051 CTTACGAACG TTGTAATTGC TAAAAATGCC GATGTCCTCA ATAATTGGGG  
 8101 TATTACGAGT TTTGAATGG CACCACAGTA CCGTTCAGT GGGGACCATA  
 8151 CATTCTTGGA TTCAACGATT GATAATGGTT ATGCCCTCAC TGATCGCTAT  
 8201 GACTTAGGTT TCAATACACC AACAAAGTAT GGCACTGATG GTGATTTGCG  
 8251 TGCAACGATT CAAGCGCTAC ATCATGCTAA TATGCAAGTT ATGGCTGACG  
 8301 TTGTTGATAA CCAGGTCTAT AACTTACCTG GTAAAGAACT TGTTCAGCA  
 8351 ACACGAGCAG GTGTTTATGG TAATGACGAC GCCACGGGCT TTGGAACGCA  
 8401 ACTCTATGTG ACTAACTCCG TTGGTGGTGG TCAATACCAA GAGAAATATG  
 8451 CTGGACAATA CTTAGAAGCT CTGAAAGCAA AGTATCCAGA CCTCTTTGAG  
 8501 GGTAAGGCCT ATGATTATTG GTATAAGAAC TATGCAATG ATGGGTCAAA  
 8551 TCCTTACTAT ACATTGTAC ACGGTGACCG TGAATCTATC CCAGCAGATG  
 8601 TTGCTATTAA GCAATGGTCA GCTAAGTATA TGAACGGCAC GAACGTTTTG  
 8651 GGCAATGGTA TGGGTTATGT ATTGAAGGAT TGGCATAATG GTCAATATTT  
 8701 CAAGCTTGAT GGTGATAAAT CAACATTACC TCAAATTTAA TTTATTTTGA  
 8751 TAGGGAACGA TTATCTTATC AAATTGTAGT GACAAAAGTC GCAGATATTG  
 8801 AATCCAATAT CTGCGACTTT TCGTCTGTAA AGCTATGCTA TAATAACGTT  
 8851 ATGACAAAAG AAAATTATTT TAAAGTTGGC ACAATTGTCA ACACCCACGG  
 8901 TATTCGTGGC GAAGTGAAGA TTATGGATAT C

# LISTE DE SEQUENCES

<110> Centre National de la Recherche Scientifique  
Institut National des Sciences Appliquées de Toulou

<120> MOLECULE D'ACIDES NUCLEIQUES CODANT UNE  
DEXTRANE-SACCHARASE CATALYSANT LA SYNTHÈSE DE DEXTRANE  
PORTANT DES RAMIFICATIONS DE TYPE ALPHA-1,2 OSIDIQUES

<130> B4787 (INPI) CNRS/INSA TOULOUSE

<140> 0103631

<141> 2001-03-16

<160> 17

<170> PatentIn Ver. 2,1

<210> 1

<211> 855

<212> PRT

<213> Acides aminés (domaine catalytique n° 2)

<400> 1

Asp Met Ser Thr Asn Ala Phe Ser Thr Lys Asn Val Ala Phe Asn His  
1 5 10 15

Asp Ser Ser Ser Phe Asp His Thr Val Asp Gly Phe Leu Thr Ala Asp  
20 25 30

Thr Trp Tyr Arg Pro Lys Ser Ile Leu Ala Asn Gly Thr Thr Trp Arg  
35 40 45

Asp Ser Thr Asp Lys Asp Met Arg Pro Leu Ile Thr Val Trp Trp Pro  
50 55 60

Asn Lys Asn Val Gln Val Asn Tyr Leu Asn Phe Met Lys Ala Asn Gly  
65 70 75 80

Leu Leu Thr Thr Ala Ala Gln Tyr Thr Leu His Ser Asp Gln Tyr Asp  
85 90 95

Leu Asn Gln Ala Ala Gln Asp Val Gln Val Ala Ile Glu Arg Arg Ile  
100 105 110

Ala Ser Glu His Gly Thr Asp Trp Leu Gln Lys Leu Leu Phe Glu Ser  
115 120 125

Gln Asn Asn Asn Pro Ser Phe Val Lys Gln Gln Phe Ile Trp Asn Lys  
130 135 140

Asp Ser Glu Tyr His Gly Gly Gly Asp Ala Trp Phe Gln Gly Gly Tyr  
145 150 155 160

Leu Lys Tyr Gly Asn Asn Pro Leu Thr Pro Thr Thr Asn Ser Asp Tyr  
165 170 175

Arg Gln Pro Gly Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp  
180 185 190

Asn Ser Asn Pro Val Val Gln Ala Glu Asn Leu Asn Trp Leu His Tyr

195

200

205

Leu Met Asn Phe Gly Thr Ile Thr Ala Gly Gln Asp Asp Ala Asn Phe  
 210 215 220  
 Asp Ser Ile Arg Ile Asp Ala Val Asp Phe Ile His Asn Asp Thr Ile  
 225 230 235 240  
 Gln Arg Thr Tyr Asp Tyr Leu Arg Asp Ala Tyr Gln Val Gln Gln Ser  
 245 250 255  
 Glu Ala Lys Ala Asn Gln His Ile Ser Leu Val Glu Ala Gly Leu Asp  
 260 265 270  
 Ala Gly Thr Ser Thr Ile His Asn Asp Ala Leu Ile Glu Ser Asn Leu  
 275 280 285  
 Arg Glu Ala Ala Thr Leu Ser Leu Thr Asn Glu Pro Gly Lys Asn Lys  
 290 295 300  
 Pro Leu Thr Asn Met Leu Gln Asp Val Asp Gly Gly Thr Leu Ile Thr  
 305 310 315 320  
 Asp His Thr Gln Asn Ser Thr Glu Asn Gln Ala Thr Pro Asn Tyr Ser  
 325 330 335  
 Ile Ile His Ala His Asp Lys Gly Val Cln Glu Lys Val Gly Ala Ala  
 340 345 350  
 Ile Thr Asp Ala Thr Gly Ala Asp Trp Thr Asn Phe Thr Asp Glu Gln  
 355 360 365  
 Leu Lys Ala Gly Leu Glu Leu Phe Tyr Lys Asp Gln Arg Ala Thr Asn  
 370 375 380  
 Lys Lys Tyr Asn Ser Tyr Asn Ile Pro Ser Ile Tyr Ala Leu Met Leu  
 385 390 395 400  
 Thr Asn Lys Asp Thr Val Pro Arg Met Tyr Tyr Gly Asp Met Tyr Gln  
 405 410 415  
 Asp Asp Gly Gln Tyr Met Ala Asn Lys Ser Ile Tyr Tyr Asp Ala Leu  
 420 425 430  
 Val Ser Leu Met Thr Ala Arg Lys Ser Tyr Val Ser Gly Gly Gln Thr  
 435 440 445  
 Met Ser Val Asp Asn His Gly Leu Leu Lys Ser Val Arg Phe Gly Lys  
 450 455 460  
 Asp Ala Met Thr Ala Asn Asp Leu Gly Thr Ser Ala Thr Arg Thr Glu  
 465 470 475 480  
 Gly Leu Gly Val Ile Ile Gly Asn Asp Pro Lys Leu Gln Leu Asn Asp  
 485 490 495  
 Ser Asp Lys Val Thr Leu Asp Met Gly Ala Ala His Lys Asn Gln Lys  
 500 505 510  
 Tyr Arg Ala Val Ile Leu Thr Thr Arg Asp Gly Leu Ala Thr Phe Asn  
 515 520 525

Ser Asp Gln Ala Pro Thr Ala Trp Thr Asn Asp Gln Gly Thr Leu Thr  
 530 535 540  
 Phe Ser Asn Gln Glu Ile Asn Gly Gln Asp Asn Thr Gln Ile Arg Gly  
 545 550 555 560  
 Val Ala Asn Pro Gln Val Ser Gly Tyr Leu Ala Val Trp Val Pro Val  
 565 570 575  
 Gly Ala Ser Asp Asn Gln Asp Ala Arg Thr Ala Ala Thr Thr Thr Glu  
 580 585 590  
 Asn His Asp Gly Lys Val Leu His Ser Asn Ala Ala Leu Asp Ser Asn  
 595 600 605  
 Leu Ile Tyr Glu Gly Phe Ser Asn Phe Gln Pro Lys Ala Thr Thr His  
 610 615 620  
 Asp Glu Leu Thr Asn Val Val Ile Ala Lys Asn Ala Asp Val Phe Asn  
 625 630 635 640  
 Asn Trp Gly Ile Thr Ser Phe Glu Met Ala Pro Gln Tyr Arg Ser Ser  
 645 650 655  
 Gly Asp His Thr Phe Leu Asp Ser Thr Ile Asp Asn Gly Tyr Ala Phe  
 660 665 670  
 Thr Asp Arg Tyr Asp Leu Gly Phe Asn Thr Pro Thr Lys Tyr Gly Thr  
 675 680 685  
 Asp Gly Asp Leu Arg Ala Thr Ile Gln Ala Leu His His Ala Asn Met  
 690 695 700  
 Gln Val Met Ala Asp Val Val Asp Asn Gln Val Tyr Asn Leu Pro Gly  
 705 710 715 720  
 Lys Glu Val Val Ser Ala Thr Arg Ala Gly Val Tyr Gly Asn Asp Asp  
 725 730 735  
 Ala Thr Gly Phe Gly Thr Gln Leu Tyr Val Thr Asn Ser Val Gly Gly  
 740 745 750  
 Gly Gln Tyr Gln Glu Lys Tyr Ala Gly Gln Tyr Leu Glu Ala Leu Lys  
 755 760 765  
 Ala Lys Tyr Pro Asp Leu Phe Glu Gly Lys Ala Tyr Asp Tyr Trp Tyr  
 770 775 780  
 Lys Asn Tyr Ala Asn Asp Gly Ser Asn Pro Tyr Tyr Thr Leu Ser His  
 785 790 795 800  
 Gly Asp Arg Glu Ser Ile Pro Ala Asp Val Ala Ile Lys Gln Trp Ser  
 805 810 815  
 Ala Lys Tyr Met Asn Gly Thr Asn Val Leu Gly Asn Gly Met Gly Tyr  
 820 825 830  
 Val Leu Lys Asp Trp His Asn Gly Gln Tyr Phe Lys Leu Asp Gly Asp  
 835 840 845

Lys Ser Thr Leu Pro Gln Ile  
850 855

<210> 2

<211> 2635

<212> PRT

<213> Acides aminés (protéine complète DSR-D)

<400> 2

Met Arg Asp Met Arg Val Ile Cys Asp Arg Lys Lys Leu Tyr Lys Ser  
1 5 10 15

Gly Lys Val Leu Val Thr Ala Gly Ile Phe Ala Leu Met Met Phe Gly  
20 25 30

Val Thr Thr Ala Ser Val Ser Ala Asn Thr Ile Ala Val Asp Thr Asn  
35 40 45

His Ser Arg Thr Ser Ala Gln Ile Asn Lys Ser Ala Val Asp Lys Val  
50 55 60

Asn Asp Asp Lys Thr Thr Leu Gly Ala Ala Lys Val Val Ala Val Ala  
65 70 75 80

Thr Thr Pro Ala Thr Pro Val Ala Asp Lys Thr Val Ser Ala Pro Ala  
85 90 95

Ala Asp Lys Ala Val Asp Thr Thr Ser Ser Thr Thr Pro Ala Thr Asp  
100 105 110

Lys Ala Val Asp Thr Thr Pro Thr Thr Pro Ala Ala Asp Lys Ala Val  
115 120 125

Asp Thr Thr Pro Thr Thr Pro Ala Ala Asp Lys Ala Val Asp Thr Thr  
130 135 140

Pro Thr Thr Pro Ala Ala Asn Lys Ala Val Asp Thr Thr Pro Ala Thr  
145 150 155 160

Ala Ala Thr Asp Lys Ala Val Ala Thr Pro Ala Thr Pro Ala Ala Asp  
165 170 175

Lys Leu Ala Asn Thr Thr Pro Ala Thr Asp Lys Ala Val Ala Thr Thr  
180 185 190

Pro Ala Thr Pro Val Ala Asn Lys Ala Ala Asp Thr Ser Ser Ile His  
195 200 205

Asp Gln Pro Leu Asp Thr Asn Val Pro Thr Asp Lys Ser Ala Asn Leu  
210 215 220

Val Ser Thr Thr Gln Lys Ser Thr Asp Asn Gln Gln Val Lys Ser Thr  
225 230 235 240

Glu Thr Ser His Leu Gln Glu Ile Asn Gly Lys Thr Tyr Phe Leu Asp  
245 250 255

Asp Asn Gly Gln Val Lys Lys Asn Phe Thr Ala Ile Ile Asp Gly Lys



Leu Lys Tyr Ala Leu Asn Met Pro Leu Ala Ala Gln Ser Gly Leu Glu  
 595 600 605  
 Pro Leu Ile Asn Thr Ser Leu Val Lys Arg Gly Lys Asp Ala Thr Glu  
 610 615 620  
 Asn Glu Ala Gln Pro Asn Tyr Ala Phe Ile Arg Ala His Asp Ser Glu  
 625 630 635 640  
 Val Gln Thr Val Ile Ala Gln Ile Ile Lys Asp Lys Ile Asn Thr Lys  
 645 650 655  
 Ser Asp Gly Leu Thr Val Thr Pro Asp Glu Ile Lys Gln Ala Phe Thr  
 660 665 670  
 Ile Tyr Asn Ala Asp Glu Leu Lys Ala Asp Lys Glu Tyr Thr Ala Tyr  
 675 680 685  
 Asn Ile Pro Ala Ser Tyr Ala Val Leu Leu Thr Asn Lys Asp Thr Val  
 690 695 700  
 Pro Arg Val Tyr Tyr Gly Asp Leu Phe Ser Asp Asp Gly Gln Tyr Met  
 705 710 715 720  
 Ser Gln Lys Ser Pro Tyr Tyr Asp Ala Ile Thr Ser Leu Leu Lys Ser  
 725 730 735  
 Arg Ile Lys Tyr Val Ala Gly Gly Gln Ser Met Asn Met Thr Tyr Leu  
 740 745 750  
 His Gln Cys Phe Asp Pro Ala Lys Asn Glu Thr Lys Pro Gln Gly Val  
 755 760 765  
 Leu Thr Ser Val Arg Tyr Gly Lys Gly Ala Met Thr Ala Asp Asp Leu  
 770 775 780  
 Gly Asn Ser Asp Thr Arg Gln Gln Gly Ile Gly Leu Val Ile Asn Asn  
 785 790 795 800  
 Lys Pro Phe Leu Asn Leu Asn Asp Asp Glu Gln Ile Val Leu Asn Met  
 805 810 815  
 Gly Ala Ala His Lys Asn Gln Ala Tyr Arg Pro Leu Met Leu Thr Thr  
 820 825 830  
 Lys Ser Gly Leu Gln Ile Tyr Asp Lys Asp Ala Gly Ala Pro Val Val  
 835 840 845  
 Tyr Thr Asn Asp Ala Gly Gln Leu Ile Phe Lys Ser Asp Met Val Tyr  
 850 855 860  
 Gly Val Ser Asn Pro Cln Val Ser Gly Tyr Phe Ala Ala Trp Val Pro  
 865 870 875 880  
 Val Gly Ala Ser Asp Ser Gln Asp Ala Arg Thr Gln Ser Ser Gln Ser  
 885 890 895  
 Glu Thr Lys Asp Gly Asp Val Tyr His Ser Asn Ala Ala Leu Asp Ser  
 900 905 910

Asn Val Ile Tyr Glu Gly Phe Ser Asn Phe Gln Ala Met Pro Glu Lys  
 915 920 925  
 Asn Asp Asp Phe Thr Asn Val Lys Ile Ala Gln Asn Ala Lys Leu Phe  
 930 935 940  
 Lys Asp Leu Gly Ile Thr Ser Phe Glu Leu Ala Pro Gln Tyr Arg Ser  
 945 950 955 960  
 Ser Thr Asp Asn Ser Phe Leu Asp Ser Val Ile Gln Asn Gly Tyr Ala  
 965 970 975  
 Phe Thr Asp Arg Tyr Asp Val Gly Tyr Asn Thr Pro Thr Lys Tyr Gly  
 980 985 990  
 Thr Val Asp Gln Leu Leu Asp Ser Leu Arg Ala Leu His Ala Gln Gly  
 995 1000 1005  
 Ile Gln Ala Ile Asn Asp Trp Val Pro Asp Gln Ile Tyr Asn Leu Pro  
 1010 1015 1020  
 Gly Glu Gln Ile Val Thr Ala Val Arg Thr Asn Gly Ser Gly Lys Tyr  
 1025 1030 1035 1040  
 Asp Tyr Asp Ser Val Ile Asn Asn Thr Leu Tyr Asp Ser Arg Thr Val  
 1045 1050 1055  
 Gly Gly Gly Glu Tyr Gln Glu Lys Phe Gly Gly Leu Phe Leu Asp Gln  
 1060 1065 1070  
 Leu Lys Lys Asp Tyr Pro Ser Leu Phe Glu Thr Lys Gln Ile Ser Thr  
 1075 1080 1085  
 Asn Gln Pro Met Asn Pro Asp Val Lys Ile Lys Glu Trp Ser Ala Lys  
 1090 1095 1100  
 Tyr Phe Asn Gly Ser Asn Ile Gln Gly Arg Gly Ala Trp Tyr Val Leu  
 1105 1110 1115 1120  
 Lys Asp Trp Ala Thr Asn Gln Tyr Phe Asn Val Ser Ser Asp Asn Gly  
 1125 1130 1135  
 Phe Leu Pro Lys Gln Leu Leu Gly Glu Lys Thr Ser Thr Gly Phe Ile  
 1140 1145 1150  
 Thr Glu Asn Gly Lys Thr Ser Phe Tyr Ser Thr Ser Gly Tyr Gln Ala  
 1155 1160 1165  
 Lys Asp Thr Phe Ile Gln Asp Gly Thr Asn Trp Tyr Tyr Phe Asp Asn  
 1170 1175 1180  
 Ala Gly Tyr Met Leu Thr Gly Lys Gln Asn Ile His Asp Lys Asn Tyr  
 1185 1190 1195 1200  
 Tyr Phe Leu Pro Asn Gly Val Glu Leu Gln Asp Ala Tyr Leu Phe Asp  
 1205 1210 1215  
 Gly Asn Gln Glu Phe Tyr Tyr Asn Lys Ala Gly Glu Gln Val Met Asn  
 1220 1225 1230



Gln Tyr Tyr Gln Asp Ser Gln Asn Gln Trp His Tyr Phe Phe Glu Asn  
 1235 1240 1245  
 Gly Arg Met Ala Ile Gly Leu Thr Glu Val Pro Asn Ala Asp Gly Thr  
 1250 1255 1260  
 His Val Thr Gln Tyr Phe Asp Ala Asn Gly Val Gln Ile Lys Gly Thr  
 1265 1270 1275 1280  
 Ala Ile Lys Asp Gln Asn Asn Gln Leu Arg Tyr Phe Asp Glu Ala Thr  
 1285 1290 1295  
 Gly Asn Met Val Val Asn Ser Trp Gly Gln Leu Ala Asp Lys Ser Trp  
 1300 1305 1310  
 Leu Tyr Leu Asn Ala Gln Gly Val Ala Val Thr Gly Asn Gln Lys Ile  
 1315 1320 1325  
 Asp Gly Glu Glu Tyr Tyr Phe Asn Ala Asp Gly Lys Gln Val Lys Gly  
 1330 1335 1340  
 Asn Ala Ile Ile Asp Asn Asn Gly Asp Gln Arg Tyr Tyr Asp Gly Asp  
 1345 1350 1355 1360  
 Lys Gly Val Met Val Val Asn Ser Trp Gly Glu Leu Pro Asp Gly Ser  
 1365 1370 1375  
 Trp Leu Tyr Leu Asn Asp Lys Gly Ile Ala Val Thr Gly Arg Gln Val  
 1380 1385 1390  
 Ile Asn Asn Gln Val Asn Phe Phe Gly Asn Asp Gly Lys Gln Ile Lys  
 1395 1400 1405  
 Asp Ala Phe Lys Leu Leu Ser Asp Gly Ser Trp Val Tyr Leu Asp Asp  
 1410 1415 1420  
 Lys Gly Leu Ile Thr Thr Gly Ala Lys Val Ile Asn Gly Leu Asn Met  
 1425 1430 1435 1440  
 Phe Phe Asp Lys Asp Gly His Gln Ile Lys Gly Asp Ala Ser Thr Asp  
 1445 1450 1455  
 Ala Asn Gly Lys Arg His Tyr Tyr Asp Lys Asn Asp Gly His Leu Val  
 1460 1465 1470  
 Thr Asn Ser Trp Gly Glu Leu Pro Asp Gly Ser Trp Leu Tyr Leu Glu  
 1475 1480 1485  
 Glu Gln Gly Asp Ala Val Thr Gly Gln Arg Val Ile Asp Gly Lys Thr  
 1490 1495 1500  
 Arg Tyr Phe Asp Glu Asp Gly Lys Gln Ile Lys Asn Ser Leu Lys Thr  
 1505 1510 1515 1520  
 Leu Ala Asn Gly Asp Lys Ile Tyr Leu Asp Gly Asp Gly Val Ala Ala  
 1525 1530 1535  
 Thr Gly Leu Gln His Val Gly Asp Lys Ile Met Tyr Phe Asp Glu Asp  
 1540 1545 1550

Gly Lys Gln Val Val Gly Lys Phe Val Ser Ala Lys Asp Gly Ser Trp  
 1555 1560 1565  
 Tyr Tyr Leu Asn Gln Asp Gly Val Ala Ala Val Gly Pro Ser Ser Ile  
 1570 1575 1580  
 Asn Gly Gln Ser Leu Tyr Phe Asp Gln Asp Gly Lys Gln Val Lys Tyr  
 1585 1590 1595 1600  
 Asn Glu Val Arg Asn Ser Asp Gly Thr Thr Asn Tyr Tyr Thr Gly Leu  
 1605 1610 1615  
 Thr Gly Glu Lys Leu Thr Gln Asp Phe Gly Glu Leu Pro Asp Gly Ser  
 1620 1625 1630  
 Trp Ile Tyr Leu Asp Ala Gln Gly His Thr Val Thr Gly Ala Gln Ile  
 1635 1640 1645  
 Ile Asn Gly Gln Asn Leu Tyr Phe Lys Ala Asp Gly Gln Gln Val Lys  
 1650 1655 1660  
 Gly His Ala Tyr Thr Asp Gln Leu Gly His Met Arg Phe Tyr Asp Pro  
 1665 1670 1675 1680  
 Asp Ser Gly Asp Met Leu Ser Asn Arg Phe Glu Gln Ile Thr Pro Gly  
 1685 1690 1695  
 Val Trp Ala Tyr Phe Gly Ala Asp Gly Val Ala Ile Thr Gly Gln His  
 1700 1705 1710  
 Asp Ile Asn Gly Gln Lys Leu Phe Phe Asp Glu Thr Gly Tyr Gln Val  
 1715 1720 1725  
 Lys Gly Ser Gln Arg Thr Ile Asp Gly Thr Leu Tyr Ser Phe Asp Ser  
 1730 1735 1740  
 Gln Thr Gly Asn Gln Lys Arg Val Gln Thr Thr Leu Leu Pro Gln Ala  
 1745 1750 1755 1760  
 Gly His Tyr Ile Thr Lys Asn Gly Asn Asp Trp Gln Tyr Asp Thr Asn  
 1765 1770 1775  
 Gly Glu Leu Ala Lys Gly Leu Arg Gln Asp Ser Asn Gly Lys Leu Arg  
 1780 1785 1790  
 Tyr Phe Asp Leu Thr Thr Gly Ile Gln Ala Lys Gly Gln Phe Val Thr  
 1795 1800 1805  
 Ile Gly Gln Glu Thr Tyr Tyr Phe Ser Lys Asp His Gly Asp Ala Gln  
 1810 1815 1820  
 Leu Leu Pro Met Val Thr Glu Gly His Tyr Gly Thr Ile Thr Leu Lys  
 1825 1830 1835 1840  
 Gln Gly Gln Asp Thr Lys Thr Ala Trp Val Tyr Arg Asp Gln Asn Asn  
 1845 1850 1855  
 Thr Ile Leu Lys Gly Leu Gln Asn Ile Asn Gly Thr Leu Gln Phe Phe  
 1860 1865 1870

Asp Pro Tyr Thr Gly Glu Gln Leu Lys Gly Gly Val Ala Lys Tyr Asp  
 1875 1880 1885  
 Asp Lys Leu Phe Tyr Phe Glu Ser Gly Lys Gly Asn Leu Val Ser Thr  
 1890 1895 1900  
 Val Ala Gly Asp Tyr Gln Asp Gly His Tyr Ile Ser Gln Asp Gly Gln  
 1905 1910 1915 1920  
 Thr Arg Tyr Ala Asp Lys Gln Asn Gln Leu Val Lys Gly Leu Val Thr  
 1925 1930 1935  
 Val Asn Gly Ala Leu Gln Tyr Phe Asp Asn Ala Thr Gly Asn Gln Ile  
 1940 1945 1950  
 Lys Asn Gln Gln Val Ile Val Asp Gly Lys Thr Tyr Tyr Phe Asp Asp  
 1955 1960 1965  
 Lys Gly Asn Gly Glu Tyr Leu Phe Thr Asn Thr Leu Asp Met Ser Thr  
 1970 1975 1980  
 Asn Ala Phe Ser Thr Lys Asn Val Ala Phe Asn His Asp Ser Ser Ser  
 1985 1990 1995 2000  
 Phe Asp His Thr Val Asp Gly Phe Leu Thr Ala Asp Thr Trp Tyr Arg  
 2005 2010 2015  
 Pro Lys Ser Ile Leu Ala Asn Gly Thr Thr Trp Arg Asp Ser Thr Asp  
 2020 2025 2030  
 Lys Asp Met Arg Pro Leu Ile Thr Val Trp Trp Pro Asn Lys Asn Val  
 2035 2040 2045  
 Gln Val Asn Tyr Leu Asn Phe Met Lys Ala Asn Gly Leu Leu Thr Thr  
 2050 2055 2060  
 Ala Ala Gln Tyr Thr Leu His Ser Asp Gln Tyr Asp Leu Asn Gln Ala  
 2065 2070 2075 2080  
 Ala Gln Asp Val Gln Val Ala Ile Glu Arg Arg Ile Ala Ser Glu His  
 2085 2090 2095  
 Gly Trp Asp Trp Leu Gln Lys Leu Leu Phe Glu Ser Gln Asn Asn Asn  
 2100 2105 2110  
 Pro Ser Phe Val Lys Gln Gln Phe Ile Trp Asn Lys Asp Ser Glu Tyr  
 2115 2120 2125  
 His Gly Gly Gly Asp Ala Trp Phe Gln Gly Gly Tyr Leu Lys Tyr Gly  
 2130 2135 2140  
 Asn Asn Pro Leu Thr Pro Thr Thr Asn Ser Asp Tyr Arg Gln Pro Gly  
 2145 2150 2155 2160  
 Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp Asn Ser Asn Pro  
 2165 2170 2175  
 Val Val Gln Ala Glu Asn Leu Asn Trp Leu His Tyr Leu Met Asn Phe  
 2180 2185 2190

Gly Thr Ile Thr Ala Gly Gln Asp Asp Ala Asn Phe Asp Ser Ile Arg  
 2195 2200 2205  
 Ile Asp Ala Val Asp Phe Ile His Asn Asp Thr Ile Gln Arg Thr Tyr  
 2210 2215 2220  
 Asp Tyr Leu Arg Asp Ala Tyr Gln Val Gln Gln Ser Glu Ala Lys Ala  
 2225 2230 2235 2240  
 Asn Gln His Ile Ser Leu Val Glu Ala Gly Leu Asp Ala Gly Thr Ser  
 2245 2250 2255  
 Thr Ile His Asn Asp Ala Leu Ile Glu Ser Asn Leu Arg Glu Ala Ala  
 2260 2265 2270  
 Thr Leu Ser Leu Thr Asn Glu Pro Gly Lys Asn Lys Pro Leu Thr Asn  
 2275 2280 2285  
 Met Leu Gln Asp Val Asp Gly Gly Thr Leu Ile Thr Asp His Thr Gln  
 2290 2295 2300  
 Asn Ser Thr Glu Asn Gln Ala Thr Pro Asn Tyr Ser Ile Ile His Ala  
 2305 2310 2315 2320  
 His Asp Lys Gly Val Gln Glu Lys Val Gly Ala Ala Ile Thr Asp Ala  
 2325 2330 2335  
 Thr Gly Ala Asp Trp Thr Asn Phe Thr Asp Glu Gln Leu Lys Ala Gly  
 2340 2345 2350  
 Leu Glu Leu Phe Tyr Lys Asp Gln Arg Ala Thr Asn Lys Lys Tyr Asn  
 2355 2360 2365  
 Ser Tyr Asn Ile Pro Ser Ile Tyr Ala Leu Met Leu Thr Asn Lys Asp  
 2370 2375 2380  
 Thr Val Pro Arg Met Tyr Tyr Gly Asp Met Tyr Gln Asp Asp Gly Gln  
 2385 2390 2395 2400  
 Tyr Met Ala Asn Lys Ser Ile Tyr Tyr Asp Ala Leu Val Ser Leu Met  
 2405 2410 2415  
 Thr Ala Arg Lys Ser Tyr Val Ser Gly Gly Gln Thr Met Ser Val Asp  
 2420 2425 2430  
 Asn His Gly Leu Leu Lys Ser Val Arg Phe Gly Lys Asp Ala Met Thr  
 2435 2440 2445  
 Ala Asn Asp Leu Gly Thr Ser Ala Thr Arg Thr Glu Gly Leu Gly Val  
 2450 2455 2460  
 Ile Ile Gly Asn Asp Pro Lys Leu Gln Leu Asn Asp Ser Asp Lys Val  
 2465 2470 2475 2480  
 Thr Leu Asp Met Gly Ala Ala His Lys Asn Gln Lys Tyr Arg Ala Val  
 2485 2490 2495  
 Ile Leu Thr Thr Arg Asp Gly Leu Ala Thr Phe Asn Ser Asp Gln Ala  
 2500 2505 2510

Fro Thr Ala Trp Thr Asn Asp Gln Gly Thr Leu Thr Phe Ser Asn Gln  
 2513 2520 2525  
 Glu Ile Asn Gly Gln Asp Asn Thr Gln Ile Arg Gly Val Ala Asn Pro  
 2530 2535 2540  
 Gln Val Ser Gly Tyr Leu Ala Val Trp Val Pro Val Gly Ala Ser Asp  
 2545 2550 2555 2560  
 Asn Gln Asp Ala Arg Thr Ala Ala Thr Thr Thr Glu Asn His Asp Gly  
 2563 2570 2575  
 Lys Val Leu His Ser Asn Ala Ala Leu Asp Ser Asn Leu Ile Tyr Glu  
 2580 2585 2590  
 Gly Phe Ser Asn Phe Gln Pro Lys Ala Thr Thr His Asp Glu Leu Thr  
 2595 2600 2605  
 Asn Val Val Ile Ala Lys Asn Ala Asp Val Phe Asn Asn Trp Gly Ile  
 2610 2615 2620  
 Thr Ser Phe Glu Met Ala Fro Gln Tyr Arg Ser Ser Gly Asp His Thr  
 2625 2630 2635 2640  
 Phe Leu Asp Ser Thr Ile Asp Asn Gly Tyr Ala Phe Thr Asp Arg Tyr  
 2645 2650 2655  
 Asp Leu Gly Phe Asn Thr Pro Thr Lys Tyr Gly Thr Asp Gly Asp Leu  
 2660 2665 2670  
 Arg Ala Thr Ile Gln Ala Leu His His Ala Asn Met Gln Val Met Ala  
 2675 2680 2685  
 Asp Val Val Asp Asn Gln Val Tyr Asn Leu Pro Gly Lys Glu Val Val  
 2690 2695 2700  
 Ser Ala Thr Arg Ala Gly Val Tyr Gly Asn Asp Asp Ala Thr Gly Phe  
 2705 2710 2715 2720  
 Gly Thr Gln Leu Tyr Val Thr Asn Ser Val Gly Gly Gly Gln Tyr Gln  
 2725 2730 2735  
 Glu Lys Tyr Ala Gly Gln Tyr Leu Glu Ala Leu Lys Ala Lys Tyr Pro  
 2740 2745 2750  
 Asp Leu Phe Glu Gly Lys Ala Tyr Asp Tyr Trp Tyr Lys Asn Tyr Ala  
 2755 2760 2765  
 Asn Asp Gly Ser Asn Pro Tyr Tyr Thr Leu Ser His Gly Asp Arg Glu  
 2770 2775 2780  
 Ser Ile Pro Ala Asp Val Ala Ile Lys Gln Trp Ser Ala Lys Tyr Met  
 2785 2790 2795 2800  
 Asn Gly Thr Asn Val Leu Gly Asn Gly Met Gly Tyr Val Leu Lys Asp  
 2805 2810 2815  
 Trp His Asn Gly Gln Tyr Phe Lys Leu Asp Gly Asp Lys Ser Thr Leu  
 2820 2825 2830

<210> 3  
<211> 2568  
<212> ADN  
<213> Nucléotides (domaine catalytique n° 2)

```

<400> 3
gatattgtcta ctaattgcttt ttctacacaa aatgttgcat tcaatcatga cagtagcagt. 60
ttcgaccata ctggtgatgg ctctcttgacg gcagatactt ggtatcgacc aaagtcaatt 120
ttggctaacg ggacaacttg gcgtgattcg actgataagg atatgcgacc attaatcact 180
gtttgtgggc caaataagaa tgttcaagtc aactacctca acttcatgaa agcaaattggc 240
ttgttgacaa cagcagcaca atacacacta cattcagatc aatatgattt gaaccaagct 300
gcacaagatg ttcaagtggc cattgaaagg cgcattgcgt cagagcatgg cacagactgg 360
ttacagaaat tgttgtttga atcacaaaat aataacccat catttgtgaa gcaacaattc 420
atttggaaca aggattctga atatcatggg ggtggtgatg cttggttcca aggtggttat 480
ctgaagtatg gcaataaacc actcacacca acaactaatt ctgattatcg tcaacctggt 540
aatgcatttg atttcttgct agcacaacgac gtggataatt ctaatcctgt tgtgcaagct 600
gaaaacttaa actggttaca ttacttaatg aactttggca ccatcactgc ggttcaagat 660
gacgctaatt ttgatatgat tcgtattgac gctgtcgact ttattcataa tgatacaatc 720
caacgtactt atgattatct tcgtgatgct tatcaagtgc aacaaagtga agccaaagca 780
aaccgcacac ttctcattgg tgaagctggc ttagacgcag gtacatcaac gattcataat 840
gatgcgttaa ttgagtcaaa cctccgtgaa gcagcgacat tgcgttaaac aaatgaacct 900
ggtaaaaata aaccattgac gaatatgcta caagacgttg acggcggtac gcttatcacc 960
gaccatacgc agaatagtac agaaaatcag gcgacaccaa actattcaat tattcacgcg 1020
cacgataaag gtgtgcaaga aaaagtaggt gcagccatta ctgatgctac tgggtctgat 1080
tggacgaact ttacagatga acagttaaaa gcgggattag agctattcta taaggatcag 1140
cgcgcaacaa acaaaaaagt taatagtatt aacataccaa gtatttatgc cctgatgttg 1200
acaaacaaag atactgttcc tcgtatgtat tatggggata tgtatcaaga tgacggacag 1260
tatatggcaa acaagagtat ctactatgat gccttagtgt cattaatgac ggctcgtaaa 1320
agctatgtca gcggtggtca aactatgagt gttgacaatc atggtttggt gaagagtgtc 1380
cgttttggaa aagatgcgat gacagctaatt gatttaggta catcagctac gcgtactgag 1440
ggtcttggtg tcattatttg taatgatcca aagttgcaac ttaatgatgc ggataaagtg 1500
acactggata tgggtgcagc acataaaaaat caaaagtatc gcgcagttat cttacaaca 1560
cgtgatggtt tggcaacctt taattcagat caagcaccaa cagcttggac aaacgatcaa 1620
ggaacgttaa cattctcaaa tcaagagatt aacgggcaag acaatacaca aattcgtggt 1680
gttgctaata cgcaagtttc tgggttatcta gctgtttggg tgcctgtggg tgcatcagac 1740
aatcaagatg cccgtacagc agcaaacgaca acagaaaatc atgatggtaa agtattacac 1800
tcgaatgcgg cattagattc taaccttatt tatgaaggtt tctctaactt ccaacctaa 1860
gcaacaacgc atgatgaact tacgaacggt gtaattgcta aaaatgccga tgtcttcaat 1920
aattggggta ttacgagttt tgaaatggca ccacagtacc gttcaagtgg ggaccataca 1980
ttcttggtt caacgattga taatggttat gccttcactg atcgctatga cttagggttc 2040
aatacaccaa caaagtatgg cactgatggg gatttgcgtg caacgattca agcgctacat 2100
catgctaata tgcaagttat ggctgacgtt gttgataaacc aggtctataa cttacctggt 2160
aaagaagttg ttccagcaac acgagcaggt gtttatggta atgacgacgc cacgggcttt 2220
ggaacgcaac tctatgtgac taactcogtt ggtggtggtc aataccaaga gaaatatgct 2280
ggacaatact tagaagctct gaaagcaaag tatccagacc tctttgaggg taaggcctat 2340
gattattggt ataagaacta tgcaaatgat ggttcaaacc cttactatac attgtcacac 2400
ggtgaccgtg aatctatccc agcagatggt gctattaagc aatggtcagc taagtatatg 2460
aacgcacaga acgttttggg caatggtatg ggttatgtat tgaaggattg gcataatggt 2520
caatatttca agcttgatgg tgataaatca acattaccto aaatttaa 2568

```

<210> 4  
<211> 8506  
<212> ADN  
<213> Séquence codant DSR-D

<400> 4

```
atgagagaca tgagggtaat ttgtgaccgt aaaaaattgt acaaatcggt caaagtacta 60
gtacacagccg gtatttttgc tttgatgatg tttggcggtca caactgctag tgttagtgca 120
aatacagattg agtttgacac gaatcatagc cgtacttcag cacagattaa taagagtgcc 180
gttgataagg ttaatgatga caagactact ttaggagcgg caaaagttagt ggagtaggcc 240
acaacgccag cgacacccgt agcagataaa acagtaagtg caccgcagc agataaggca 300
gtagalacua cglcatccac gacacctgca acggataagg cagtagatac aacgccaacg 360
acacctgcag cagataagyc agtagataca acgccaacga cacctgcagc agataaggca 420
gtagatacaa cgccaacgac acctgcagca aataaagcag tagatacaac gccagcgacc 480
gctgcaacag ataaggcggg agccacgcca gccacacctg cagcagataa gctagcaaat 540
acgacgcctg caacggacaa ggtagtagcc acaacgccag cgacgccggt agcaataaaa 600
gcagcagaca cgagtagtat tcatgatcaa ccattagata caaatgtgcc aactgataaa 660
tcagcaaaacc tcgtctcgac aacacaaaaa agtacggata atcaacaagt taagtctaca 720
gaaacatctc atcttcaaga aatcaacggg aaaacctatt ttcttgacga caatggtcaa 780
gttaaaaaga acctcaccgc tattattgac ggtaaaagtlc tatactttga taaaacatcc 840
ggcgaattga cgcgaatgc accgcaagtt actaagggat tagtaaatat tgataatgca 900
cataacgcgg ctcctgatct cacagctgat ttaaaaaacg gaacgacctg gacaccaaca 1020
gctaacaagt ggtagctgcc taaggacatc tcttggtggc cggataagaa tacgcaggta 1080
acagcagaag attttcgacc attgctcatg tcttggtggc acgatgttaa agtatcaaat 1140
gcttatctac aatatacgca atcagttggt gctatgactg ttcaaaagaa tatcgaatcg 1200
gatgataata tgaycacatt gactgatgct ctcaagcaag atatgaacaa actgattgat 1260
cgaattgggt tatctggaaa aactgaalgg ctcaagggta atgatcattt acaggggtgg 1320
tcacagcgaa atttgaatat tgatagttaa tcaaaagggta actcagatta cgtctgttta 1380
gcattgttat atgtgaatga aacgggccaa attactgatc caagtaacaa aggtggatat 1440
aacogtacac caaccaacca aacogggccaa tctaaccttg ttgtacaagc tgagcaattg 1500
gagatgttat tagctaata tggtgataat tctaaccttg ttgtacaagc tgagcaattg 1560
aactggcttc actacatgat gaacattggt actatagclc agaacgacct aacagctaat 1620
tttgacgggt atcgtgttga tgcggtgat aacgttgatg cggatcaagc aaacaatcat 1680
ggtagattac ttaaagctgc atacggtaact ggtaaaactg aggcacacgc aaacaatcat 1740
atttgcgtct tggagcattg ggataataat gattctgcgt acattaaagc ccacgggaat 1800
aaccuaalga caatggattl tccagccacac aataacgctt aatacgcctt gaacatgcct 1860
cttgcgcgac aaagtggcct agaaccgcta attaatacaa gtctgttaa gcgtgggaaa 1920
gatgccacag aaaaatgaagc acaaccaaac tatgccttta tccgtgcccc tgatagttaa 1980
gtgcagaccg trattgcaca aattaltaa gataaaatta acacaaaatc agacggctta 2040
actgtaacac caqatgagat taagcaagct ttcactatrt acaacgccga tgaattaaaa 2100
gcagataagg aatatacagc atacaatatt cctgcttctt acgctgtatt gttgacaaac 2160
aaggtactgt tgccacgtgt ttattatggt gatctatttt ctgatgatgg acagtatatg 2220
tcacagaagt caccatacta tgacgccatt acgtcacttt tgaaaagccg tatcaaatat 2280
gtlgtggtg gtcaaaagtat gaatatgacg taottgcatg agtgcttga tccagcaaaa 2340
aatgagacaa agccacaagg tgtcttaaca tcagtacgtt acggtaaagg tgcgatgacg 2400
gctgacgatt tgggtaatat tgacacacgt caacaaggta ttgggttggg gattaataat 2460
aagccartct tgaattttaa tgatgatgae caaattgtgc tcaatatggg tgctgctcac 2520
aaaaatcaag cttaccgacc acttatgttg aacgatgctg ctggcttcca aatttacgat 2580
aaggatgcgg ggcgcgaggt tgttttact gtatctggtt attttgctgc atgggtacca 2640
gatctggtct atggtgtcag caatccacag gtatctggtt attttgctgc atgggtacca 2700
gtcgggtcga gtgatagta agatgtctaga acacaagca gccagtcaga aactaaggat 2760
ggcgatgctc atcattcaaa tgctgcgctt gattctaagt tgatttatga aggtctctcg 2820
aatttccaaq caatgcctga asagaatgat gacttcacca acgtaaaaat tgctcaaaat 2880
gctaaattgt ttaaagattc aggyattaca agctttgaat tagcaccgca atatcgttca 2940
agtacagata atagtttttt ggattcggtt atccaaaacg gctatgcctt tactgatcga 3000
tatgatgttg gctataatlc gccacaaaaa tatgggtacg actgggtacc tgatcaaat 3060
ctaagagcat tacacgcaca aggtatlcag gctattaatg actgggtacc aggtaaagtac 3120
tataatllac ctggcgcaaa aatcgtaacc gcgttctgta caaatgggtc aggtaaagtac 3180
gattatgati cagtgartha taacacgctc tatgactcac gaaacagttg ggcgggcgaa 3240
taccaagaaa agtttgggtg cctgttctta gaccagttga aaaaagatta tctagcttg 3300
tttgaacta agcagatatc aacgaatcag cagatgaalc cggatgttwa aattaaagaa 3360
tgytctgcaa agtactttaa tggttcaaac attcaaggtc gtggcgcttg gtatgtactt 3420
aaagactggg caacaaatca atatttcaat gtgtctagt ataattgatt cttgctctaa 3480
cagttactgg gtgaaaaaac aagcaccggc tttataacag aaaatggtaa gacttctttc 3540
tactcaacaa gtggttatca agctaaagat acctttattc aagatggaac aaattggtat
```

tactttgata	atgcaggcta	tatgttgaca	ggtaaacaaa	atatccacga	taaaaattat	3600
tatttcttac	ctaattggtg	ggaacttcaa	gatgcttacc	tttttgatgg	taatcaagaa	3660
ttttactata	ataaagctgg	ggaacaagtt	atgaaccagt	attatcaaga	tagtcaaaat	3720
caatggcatt	atttctttga	aaatggctgc	atggcaattg	gcctgacaga	agttccgaac	3780
gctgatggca	cccatgttac	acaatatttt	gatgctaatt	gtgtccaaat	taaaggcaca	3840
gctataaaa	atcagaataa	tcaattacgc	tattttgatg	aggccacagg	taatatgggtg	3900
gttaattcat	ggggacagtt	agcagataag	tcttggcctt	accttaatgc	acaaggcggt	3960
gctgtgactg	gtaaccaaaa	aatgtatggt	gaagagtact	acttcaatgc	tgatggtaag	4020
caagttaaag	gcaatgcaat	catcgataat	aatggtgatc	aacgttatta	tgatgggtat	4080
aagggtgtca	tggtagttaa	ttcatggggt	gagttgccag	atggctcatg	gttatatttg	4140
aatgacaaag	gtattgctgt	qacaggccgt	caagtcatta	ataatcaagt	taatttcctt	4200
ggtaatgatg	gtaagcaaat	caaagatgcc	tttaaattat	tatccgatgg	ttcatgggtg	4260
tatttggatg	ataagggcct	gataacaact	ggagccaaag	ttatcaatgg	tctaaatatg	4320
ttttttgata	aagacgggtc	tcaaatcaaa	ggtgatgcca	gcacggatgc	caatggtaag	4380
cgccattatt	atgacaaaaa	tgatgggtcat	cttgtcacia	attcatgggg	tgagtggcca	4440
gatggttcat	ggttatatct	agaagaacaa	ggtgatgctg	ttactgggtc	acgtgtgatt	4500
gatggcaaga	cacgctattt	tgatgaagat	ggcaaacaaa	ttaaaaaatg	cctaaaaacg	4560
ctggccaatg	gcgataagat	ttatcttgat	ggtgatgggg	ttgctgcaac	aggcttacia	4620
catgtgggog	ataaaatcat	gtattttgat	gaagatggca	aacaagttgt	tggcaagttt	4680
gtatcagcaa	aagatgggtc	atgggtattac	ttaaatcagg	atgggtgtgc	cgcggttggt	4740
ccaagcagca	ttaatggaca	atcacctttac	tttgatcaag	atggtaaac	agttaaatat	4800
aatgaagttc	gtaatagtga	tggaacaacc	aactattaca	caggattaac	gggtgaaaag	4860
ttaacgcga	acttcgggtc	actaccagat	ggttcatgga	tttatcttga	tgcgcaaggt	4920
catacagtaa	ctggtgcaca	aatcattaac	ggtcaaaatc	tttactttaa	ggctgacggc	4980
cagcaagtta	aaggtcatgc	ttatactgac	caattaggtc	atatgcgttt	ttatgatcct	5040
gattcagggtg	atatgttgag	taatcgcttt	gaacaaatca	cacctgggtg	atgggcttac	5100
tttgggtgctg	atggtgtggc	cataactgga	caacatgaca	taaatgggtc	gaagctattc	5160
tttgatgaga	caggatatca	agttaaaggt	tcgcaacgta	caatagatgg	tacgttatac	5220
agcttcgatt	ctcaaactgg	taaccaaaaa	cgctacaga	caacattgtt	gccacaagca	5280
ggctactata	tcacgaaaaa	tggttaacgat	tggcagtatg	ataccaatgg	tgaactagcg	5340
aagggtctgc	gtcaagatag	caatggtaag	ttgcgttact	ttgatttgac	aaccggcata	5400
caagcgaaag	gccaatttgt	tacaattggc	caagaaactt	attactttag	taaagatcac	5460
ggggatgcgc	agttattggc	aatggtcact	gaagggcatt	acggtacaat	aacactcaag	5520
caaggccaag	acacaaaaac	agcctggggt	tacgctgac	aaaataatac	tattttgaag	5580
ggattgcgaa	atatcgaatg	cacgttgcaa	ttctttgatc	catatacagg	tgaacaactt	5640
aagggtggcg	tagcaaaagta	tgacgacaag	ctcttttact	ttgaatcagg	taaaggtaat	5700
cttgttagca	ccgtagcagg	tgactatcag	gatggtcatt	atatttccca	agatggccaa	5760
acacggttacg	gtcaagatag	aaatcagctt	gtaaaaggac	ttgttactgt	taatggggca	5820
ttacaatact	ttgataacgc	tactggtaac	caaataaaaa	atcaacaagt	tattgttgat	5880
ggcaagacgt	actattttga	cgataaaggc	aatggtgaat	acttattcac	taatacatta	5940
gatatgtcta	ctaattgctt	ttctaccaaa	aatgttgcat	tcaatcatga	cagtagcagt	6000
ttcgaccata	ctgttgatgg	cttcttgacg	gcagataact	ggtatcgacc	aaagtcaatt	6060
ttggctaacg	ggacaacttg	gcgtgattcg	actgataagg	atatgcgacc	attaatcact	6120
gtttggtggc	caaataagaa	tgttcaagtc	aactacctca	acttcatgaa	agcaaatggc	6180
ttgttgacaa	cagcagcaca	atacacacta	cattcagatc	aatatgattt	gaaccaagct	6240
gcacaagatg	ttcaagtggc	cattgaaagg	cgcatgtcgt	cagagcatgg	cacagactgg	6300
ttacagaaat	tgttgtttga	atcacaaaa	aataaaccat	catttgtgaa	gcaacaattc	6360
atttggaaaca	aggattctga	atatcatggt	ggtggtgatg	cttgggtoca	aggtggttat	6420
ctgaagtatg	gcaataaccc	actcacacca	acaactaatt	ctgattatcg	tcaacctggt	6480
aatgcatttg	atttcttgct	agccaaacgac	gtggataatt	ctaatacctgt	tgtgcaagct	6540
gaaaacttaa	actggttaca	ttacttaatg	aactttggca	ccatcactgc	gggtcaagat	6600
gacgctaatt	ttgatagtat	tcgtattgac	gctgtcgact	ttattcataa	tgatacaatc	6660
caacgtactt	atgattatct	tcgtgatgct	tatcaagtgc	aacaaaagtg	agccaaagca	6720
aaccagcaca	tttcattggt	tgaagctggc	ttagacgcag	gtacatcaac	gattcataat	6780
gatgcgttaa	ttgagtcaaa	cctccgtgaa	gcagcgacat	tgctgttaac	aatgaacct	6840
ggtaaaaaata	aaccattgac	gaatatgcta	caagacgttg	acggcggtac	gcttatcacc	6900
gaccatacgc	agaatagtac	agaaaaatcag	gcgacaccaa	actattcaat	tattcacgcg	6960
cacgataaaag	gtgtgcaaga	aaaagtaggt	gcagccatta	ctgatgctac	tggtgctgat	7020
tggaagcaact	ttacagatga	acagttaaaa	gccggattag	agctattcta	taaggatcag	7080
cgcgcaacaa	acaaaaagta	taatagttat	aacataccaa	gtatttatgc	cctgatgttg	7140
acaaacaaa	atactgttcc	tcgtatgtat	tatggggata	tgtatcaaga	tgacggacag	7200



tatatggcna	acaagagtat	ctactatgat	gccttagugt	cattaatgac	ggctcgtaaa	7260
agctatgtca	gcggtgggtca	aacctatgagt	gttgacaatc	atgggttgtt	gaagagtgtc	7320
cgttttggaa	aagatgcgal	gacagctaet	gatttaggta	carcagctac	gcgtactgag	7380
ggcttcgggt	tcattattgg	taatgatcca	aagttgcaac	ttaatgattc	ggataaagt	7440
acactggata	tgggtgcagc	acataaaaaat	caaaagtatc	gcgcagttat	cttaacaaca	7500
cgatgatggt	tggcaccctt	taattcagat	caagaccaca	cagcttggac	aaacgatcaa	7560
ggaacgttaa	caitcccaaa	ccaagagatt	aacgggcaag	acaatacaca	aattcgtggt	7620
gttgctaatc	cgcaaglttc	tgggtatcta	gctgtttggg	tgcctgtggg	tgcactagac	7680
aaccaagatg	cccgtaacgc	agcaacgaca	acagaaaatc	argatggtaa	agtattacac	7740
tcgaatgcgg	caitagattc	taaccttatt	tatgaagggt	tctctaactt	ccaacctaa	7800
gcaacaacgc	atgatgaact	tacgaacgtt	gtaattgcta	aaaatgcga	tgtcttcaat	7860
aattggggta	ttacgagttt	tgaatggca	ccacagtacc	gttcaagtgg	ggaccataca	7920
ttcttggatt	caacgattga	taattggtat	gccttcactg	atcgctatga	cttaggtttc	7980
aatacaccua	caaagtatgg	cactgatggt	gatttgcgtg	caacgattca	agcgctacat	8040
catgctaata	tgaagttat	ggctgacgtt	gttgataacc	aggtctataa	cttacctggt	8100
aaagaagctg	tttcagcaac	acgagcaggt	gtttatggla	atgacgacgc	cacgggcttt	8160
ggaacgcaac	tctatgtgac	taactccgtt	gggtggtggtc	aataccaaga	gaaatatgct	8220
ggacaatact	tagaagctct	gaaagcaaa	tatccagacc	tctttgaggg	taaggcctat	8280
gattatrggt	ataagaacta	tgcaaatgat	gggtcaaatc	cttactatac	attgtcacac	8340
ggtgaccgtg	aatctatccc	agcagatggt	gctattaagc	aatggctcagc	taagtatatg	8400
aacggcacga	acgttttggg	caatgggtatg	ggttatgat	tgaaggattg	gcataacggt	8460
caatatttca	agcttgatgg	cgataaatca	acattacctc	aaattt		8506

<210> 5  
 <211> 8931  
 <212> ADN  
 <213> Gène dsr-D

<400> 5	aataatcrgt	clccarttqct	ttcaaaataa	taataqgtta	ttattatcat	ggaacaatca	60
	atatattati	tatatattcact	atcgaatato	cttttttgc	taaatctcta	gagccgattt	120
	tttgggttat	acaanryaatt	ggtaaaaggt	aatcattttt	acaaaaccat	ggcgggtttt	180
	tattttttct	aaazuttaccg	aactagagga	agagaaaaag	agcaatagtt	gtatyagaga	240
	catgagggta	atttgtgacc	gtacaaatcg	ggcaaagtac	tagtaacagc	300	
	cggtattttt	gctttgatga	rgtttggcgt	cacaactgct	agtgttagtg	cnaatacgtat	360
	tgcagttgac	acgaatcata	gccgtacttc	agcacagatt	aataagagtg	ccgttgataa	420
	ggttaantgat	gacaagacta	cttlagygag	ggcaaaagta	glggcagtag	ccacaacgcc	480
	agcgacaccg	gtagccagata	aaacagtaag	tgcacccgca	gcagataaag	cagtagatac	540
	aacgctcatca	acgacacctg	caacgggataa	ggcagtagat	acaacgcgca	cgacacctgc	600
	aqcagataag	gcagtagata	caacgcacaac	qacacctgca	gcagataaag	cagtagatac	660
	aacqccaacg	acacctgcag	caaataaagc	agtagataca	acgccagcga	ccgtgcnaac	720
	agataaggcy	gtagccacgc	cagccacacc	tgcagcagat	aagctagcaa	atacagcgcc	780
	tgaacaggac	aaggcagtag	ccacaacgcc	agcgacgcgg	gtagcaataa	aagcagcaga	840
	cacgagtagt	altcatgata	aacctattaga	tacaaatgtg	ccaactgata	aatcagcaaa	900
	cctcgtctcg	acancacaea	aaagtacgga	caatcaacaa	gttaagtcta	caaaaaacac	960
	tcactcttcaa	gaatcaacg	gtaaaacct	tttcttgac	gacaatggtc	aagttaaaaa	1020
	gaacttcacc	gctattattg	acggtaaggt	ctctactttt	gataaaacat	ccggcggaatt	1080
	gaccgcaaat	gcaccgcaag	ttactcaggg	attagtaaat	attgataatg	cacataacgc	1140
	ggctcatgat	ctcacagcrg	ataactlcac	aatgtcgar	ggttacttaa	cagctaacag	1200
	ttggtalagt	cctaaggaca	tcttaaaaaa	cggaaacgacc	tggacaccaa	caacagcaga	1260
	agaattttoga	ncattgcaca	tgrcttggtg	gcgggataag	aalacgcagg	taycttatct	1320
	acaatatatg	caatcayttg	gtatgctacc	tqacgatgtt	aaagtalcaa	atgatgataa	1380
	catgagcaca	tgaactgatg	crgctatgac	tgttcaaaag	aatatcgaa	cgcgaattgg	1440
	tgtatctgga	aaactgatt	ggctcagca	agacatgaac	aaactgattg	attcacaggc	1500
	aaatttgaat	attgatagtg	aatcaaggg	taatgatcat	ttacaggggtg	gggcattgtt	1560
	atacgtgaat	gatgacaaaa	cucetaacyc	gaactcagat	taccgtctgt	taaaccgtac	1620
	accaaccaac	caaacggguc	aaattactga	tccaagtaaa	caagggtggat	atgagatgtt	1680
	attagctaac	gatgttgata	nttctaacc	tgttgtaaaa	gctgagcaat	tgaactggct	1740
	tcactacatg	atgaacattg	gtactatagc	tcagaacgac	ccaacagcta	attttgacgg	1800

ttatcgtgtt	gatgcggtt	ataacgttga	tgccgatctc	ttacaaattg	ctggtgatta	1860
ctttaaagct	gcatacggta	ctggtaaaac	tgaggcaaac	gcaaacaatc	atatttcgat	1920
cttggaagat	tgggataata	atgattctgc	gtacattaaa	gcccacggga	ataaccaatt	1980
gacaatggat	tttccagcac	acttggcttt	gaaatacggc	ttgaacatgc	ctcttgccgc	2040
acaaagtggc	ctagaaccgc	taattaatac	aagtcttggt	aagcgtggga	aagatgccac	2100
agaaaatgaa	gcacaaccaa	actatgcctt	tatccgtgcc	catgatagtg	aagtgcagac	2160
cgttattgca	caaattatta	aggataaaat	taacacaaaa	tcagacggct	taactgtaac	2220
accagatgag	attaagcaag	ctttcactat	ttacaacgcc	gatgaattaa	aagcagataa	2280
ggaatataca	gcatacaata	ttcctgcctc	ttacgctgta	ttgttgacaa	acaaggatac	2340
tgtgccacgt	gtttattatg	gtgatctatt	ttctgatgat	ggacagtata	tgtcacagaa	2400
gtcaccatac	tatgacgcca	ttacgtcact	tttgaaaagc	cgtatcaaat	atgttgctgg	2460
tgggtcaaagt	atgaatatga	cgtacttgca	tgagtgcctt	gatccagcaa	aaaatgagac	2520
aaagccacaa	ggtgtcttaa	catcagtagc	ttacggtaaa	ggtgcatga	cggctgacga	2580
tttgggttaa	agtgaacac	gtcaacaagg	tattggtttg	gtgattaata	ataagccatt	2640
cttgaattta	aatgatgatg	aacaaattgt	gctcaatatg	ggtgctgctc	acaaaaatca	2700
agcttaccga	ccacttatgt	tgacaacaaa	atctggtctt	caaatttacg	ataaggatgc	2760
cggagcgcca	gttgtttata	ctaacgatgc	tggccaactt	atttttaagt	cagatatggt	2820
ctatggtgtc	agcaatccac	aggatctggc	ttattttgct	gcattgggtac	cagtcgggtg	2880
gagtgatagt	caagatgcta	gaacacaaag	cagccagtca	gaaactaagg	atggcgatgt	2940
ctatcattca	aatgctgcgc	ttgattctaa	tgtgatttat	gaaggcttct	cgaatttcca	3000
agcaatgcct	gaaaagaatg	atgacttcac	caacgtaaaa	attgctcaaa	atgctaaaatt	3060
gtttaaagat	ttagggatta	caagctttga	attagcaccg	caatatcgtt	caagtacaga	3120
taatagtatt	ttgggttcgg	ttatccaaaa	cggctatgcc	tttactgac	gatatgatgt	3180
tggctataat	acgccaaaca	aatgtgtac	agttgatcaa	ctctagata	gtctaagagc	3240
attacacgca	caaggtattc	aggctattaa	tgactgggta	cctgatcaaa	tttataattt	3300
acctggcgaa	caaatcgta	cgcgagttcg	tacaaatggt	tcaggtaagt	acgattatga	3360
ttcagtgatt	aataacacgc	tctatgatcc	acgaacagtt	ggggcgcgcg	aataccaaga	3420
aaagtttggg	ggcctgttct	tgcacaggtt	gaaaaaagat	tatcctagct	tgtttgaaac	3480
taagcagata	tcaacgaatc	agccgatgaa	tccggatggt	aaaattaaag	aatgggtctgc	3540
aaagtacttt	aatggttcaa	acattcaagg	tcgtggcgct	tggtagtac	ttaaagactg	3600
ggcaacaaat	caatatttca	atgtgtctag	tgataatgga	ttcttgccca	aacagttact	3660
gggtgaaaaa	acaagcaccg	gctttataac	agaaaatggt	aagacttctt	tctactcaac	3720
aagtggttat	caagctaaag	atacctttat	tcaagatgga	acaaattggt	attactttga	3780
taatgcaggc	tatatgttga	caggtaaaac	aaatatccac	gataaaaatt	attatttctt	3840
acctaatggg	gtggaaacttc	aagatgctta	cctttttgat	ggtaatcaag	aattttacta	3900
taataaagct	ggggaacaag	ttatgaacca	gtattatcaa	gatagtcaaa	atcaatggca	3960
ttatttcttt	gaaaatggtc	gcattggcaat	tggcctgaca	gaagttccga	acgctgatgg	4020
cacctatggt	acacaatatt	tgtgtgctaa	tttaaaggca	acagtataaa	cagctataaa	4080
agatcagaat	aatcaattac	gctattttga	tgaggccaca	ggtaatatgg	tggttaattc	4140
atggggacag	ttagcagata	agtccttgct	ttaccttaat	gcacaaggcg	ttgctgtgac	4200
tggtaaccac	aaaatttgatg	tggaagagta	ctacttcaat	gctgatggtg	agcaagttaa	4260
aggcaatgca	atcatcgata	ataatggtga	tcaacgttat	tatgatggtg	ataagggtgt	4320
catggtagtt	aattcatggg	gtgagttgcc	agatgggtca	tggttatatt	tgaatgacaa	4380
aggtagttgt	gtaacaggcc	gtcaagtcac	taataatcaa	gttaatttct	ttggtaatga	4440
tggttaagcaa	atcaaaagatg	cctttaaatt	attatccgat	ggttcctggg	tgtatttga	4500
tgataagggc	ctgataacaa	ctggagccaa	agttatcaat	ggtctaaata	tgttttttga	4560
taaagacggt	catcaaatca	aagggtgatgc	cagcacggat	gccaatggta	agcgccatta	4620
ttatgacaaa	aatgatggtc	atcttgtcac	aaattcatgg	ggtgagttgc	cagatgggtc	4680
atggttatat	ctagaagaac	aagggtgatgc	tgttactggt	caacgtgtga	ttgatggcaa	4740
gacacgctat	tttgatgaag	atggcaaaaca	aattaaaaat	agcctaaaaa	cgtggtgcaa	4800
tggcgataag	attttatctg	atgggtgatgg	ggttgctgca	acaggcttac	aacatgtggg	4860
cgataaaaatc	atgtattttg	atgaagatgg	caaacaagtt	gttggcaagt	ttgtatcagc	4920
aaaagatggg	tcatggtatt	acttaaatca	ggatgggtgt	gccgcgggtg	gtccaagcag	4980
cattaatgga	caatcacttt	actttgatca	agatggtaaa	caagttaaat	ataatgaagt	5040
tcgtaatagt	gatggaaaca	ccaactatta	cacaggatta	acgggtgaaa	agttaacgca	5100
agacttcggt	gaactaccag	atggttcatg	gatttatctt	gatgcgcaag	gtcatacagt	5160
aactgggtgca	caaatcatta	acgggtcaaaa	tctttacttt	aaggctgacg	gccagcaagt	5220
taaaggtcat	gcttatactg	accaattagg	tcatatgcgt	ttttatgac	ctgattcagg	5280
tgatatgttg	agtaatcgct	ttgaacaaat	cacacctggt	gtatgggctt	actttggtgc	5340
tgatgggtg	gccataactg	gacaacatga	cataaatggt	cagaagctat	tctttgatga	5400
gacaggatat	caagttaaag	gttcgcaacg	tacaatagat	ggtacgttat	acagcttcga	5460

ttcrcaaaact	ggtaaccaa	aacgcgtaca	gacaacattg	trgccacaag	caggtcacta	5520
taccacgaaa	aatggtaacg	attggcagta	tgataccaat	ggatgaactag	cgaaggggtct	5580
gcglcaagat	agcaatggla	agtfgcglta	ctttgatttg	acaaccggca	tacaagcgaa	5640
aggccaattt	gttacaattg	gccaaagaa	ttattacttt	agaaaagatc	acggggatgc	5700
gcagltattg	ccaatggta	ctgaagggca	ttacggta	ataaactca	agcaaggtca	5760
agacaccaa	acaagcctggg	ltacacgtga	tcaaaataat	actattttga	agggattgca	5820
aaatalcaat	ggcagcgtgc	aatlctttga	tccatataca	ggatgaacaac	ttaggggtgg	5880
cgtagcaaa	tatgacgaca	agcctttta	ctttgaatca	ggtaaaggta	atcttggttag	5940
caccgtagca	ggtagactatc	aggatgtgca	ttatatttcc	caagatggcc	aaacacgtta	6000
cgcagataag	caaaatcagc	ttgtaaaggg	acttgttact	gttaatgggg	cattacaata	6060
ctttgataac	gctactggtta	accaaataaa	aatcaacaa	gttatlggtg	atggcaagac	6120
gtactatttc	gacgataaa	gcaatgggtga	atacttattc	actaatatcat	tagatatgtc	6180
tactaatgct	ttttctacca	aaaatgttgc	attcaatcat	gacagtagca	gtttcgacca	6240
tactgttgat	ggcttcttga	cggcagatag	ttggatcgca	ccaaagtcaa	ttttggctaa	6300
cgggacaact	ttggcgtgatt	cgactgataa	ggatatcgga	ccattaatca	ctgtttgggtg	6360
gccaaataag	aatgttcaag	tcaactacct	caacttcatg	aaagcaaatg	gcttgttgac	6420
aacagcagca	caatacacac	tacattcaga	tcaatgat	ctgaaccaag	ctgcacaaga	6480
tggtcaagtg	gccattgaaa	ggcgcattgc	gtcagagcat	ggcacagact	ggttacagaa	6540
attgtgtgtt	gaatcacaaa	ataataacc	atcatttgtg	aagcaacaat	tcatttggaa	6600
caaggattct	gaatatcatg	gtgtgtgtga	tgcttgggtc	caaggtgggtt	atctgaagta	6660
tggaataaac	ccactcacac	caacaactaa	ttctgattat	cgtcaacctg	gtaatgcatt	6720
tgatttcttg	ctagccaacg	acgtgggataa	ttctaactct	gttgtgcaag	ctgaaaactt	6780
aaactgggta	cattacttaa	tgaacttgg	caccatcaact	gcgggtcaag	atgacgctaa	6840
ttttgatagt	attcgtattg	acgtgttgcga	ctttattcat	aatgatacaa	tccaacgtac	6900
ttatgattat	cttcgtgagt	cttatcaagt	gcaacaaagt	gaagccaaag	caaaccagca	6960
carttccattg	gttgaagctg	gcttagacgc	aggtagcatca	acgatrcata	atgatgcgtc	7020
aatcgagtca	aacctccgtg	aagcagcgac	attgtcgtta	acaaatgaac	ctggtaaaaa	7080
taaaacttg	acgaatattg	tacaagacgt	tgacggcggg	acgcttatca	cagaccatc	7140
gcagantag	acagaaaatc	aggcgacacc	aaactattca	attattcag	cgcacgataa	7200
aggtgtgcaa	gaaaaagtag	gtgcagccat	tactgatgct	actggtgctg	attggacgaa	7260
ctttacagat	gaacagttaa	aaqccggatt	agagctattc	tataaggatc	agcgcgcaac	7320
aaacaaaag	tataatagtt	ataacatacc	aagtatttat	gccctgatgt	tgacaacaa	7380
agatcctgtc	ccicgtatgt	attatgggga	tatgtatcaa	gargacggac	agtatatgyc	7440
aaacaaaggt	atctactatg	atgccttagt	gtcattaatg	acggctcgta	aaagctatgt	7500
cagcggggt	caaaactatga	glttgacaa	tcattggtttg	ttgaagagt	tcctgtttgg	7560
aaaagatgcg	atgacagcta	atgntttagg	tacatcagct	acgcgtactg	agggctttgg	7620
tgatcattart	ggtaatgatc	caaagttgca	acttaatgat	tcggataaag	tgacactgga	7680
tatgggtgca	gcacataaaa	atcaaaagta	tcgcgcagtt	atcttaacaa	cacgtgatgg	7740
tttggcaacc	cttaactcag	atcaagcacc	aacagcttgg	acaaacgac	aaggaacggt	7800
aacattctca	aatcaagaga	ttaacgggca	agacaatata	caaattcgtg	gtgttgctaa	7860
tccycaagtt	tctgggtatc	tagctgtttg	ggtagcctgtg	ggtagcatcag	acaatcaaga	7920
lgcccgta	gcagcaacga	caacgaaaaa	tcatgatggt	aaagtattac	actogaatgc	7980
ggcatttgat	tctaactta	tttatgaagg	ttlcrctaac	ttccauccta	aggcaacaa	8040
gcatgatgaa	cttacgaacg	tlgtatgtgc	taaaaatgcc	gatgrcttca	ataattgggg	8100
tattacagagt	tttgaatgg	caccacagta	ccgttcaagt	ggggaccata	cattcttgya	8160
ttcaacgatt	gataatgggt	atgccttcac	tgatcgctat	gacttaggtt	tcaatacacc	8220
aaacaaagt	ggcactgatg	gtgatttcg	tgcaacgatt	caagcgctac	atcatgctaa	8280
latgcaagtt	atggctgacg	ltgtgtgtaa	ccaggtctat	aacttacctg	gtaaagaagt	8340
tgtttcagca	acacgagcag	gtgttatgg	raatgacgac	gcacagggct	ttggaacgca	8400
actctatgtg	actaactccg	ttgggtgtgg	tcaataccaa	gagaaatag	ctggacaata	8460
cttagaagct	ctgaanagca	agtatccaga	cctctttgag	ggtaaggcct	atgattattg	8520
gtoraagac	tatgcaaatg	atgggtcaaa	tccttactat	acattgtcac	acggtgaccg	8580
gpatctalc	ccagcagatg	trgtatttaa	gcaatgggtca	gcraagata	lgaacggcac	8640
gaacgttttg	ggcaatggt	ttgggttatgt	attgaaggat	tggralzatg	grcaatattt	8700
caagcttgat	gggtgataaat	caacattacc	tcaaatlttaa	ttatlttga	tagggaaacga	8760
ttatcttacc	aaattgtatg	gacaaaagtc	gcagatattg	aatccaatct	ctgcgacttt	8820
rugctgtgaa	agctatgca	taacacggtt	atgacaaaag	aaaattattt	taaggttggc	8880
acaattgtca	acacccacgg	tattcgtggc	gaagtgaaga	ttatggatat	c	8931

<210> 6  
<211> 17  
<212> PRT  
<213> Peptides

<400> 6  
Ala Asn Trp Asn Ile Asp Ser Glu Ser Lys Gly Asn Asp His Leu Gln  
1 5 10 15

Gly

<210> 7  
<211> 24  
<212> PRT  
<213> Peptides

<400> 7  
Gly Gly Tyr Glu Met Leu Leu Ala Asn Asp Val Asp Asn Ser Asn Pro  
1 5 10 15

Val Val Gln Ala Glu Gln Leu Asn  
20

<210> 8  
<211> 21  
<212> PRT  
<213> Peptides

<400> 8  
Ala Asn Phe Asp Gly Tyr Arg Val Asp Ala Val Asp Asn Val Asp Ala  
1 5 10 15

Asp Leu Leu Gln Ile  
20

<210> 9  
<211> 12  
<212> PRT  
<213> Peptides

<400> 9  
His Ile Ser Ile Leu Glu Asp Trp Asp Asn Asn Asp  
1 5 10

<210> 10  
<211> 15  
<212> PRT  
<213> Peptides

<400> 10  
Tyr Ala Phe Ile Arg Ala His Asp Ser Glu Val Gln Thr Val Ile  
1 5 10 15

<210> 11  
<211> 8  
<212> PRT  
<213> Peptides

<400> 11  
Asp Trp Val Pro Asp Gln Ile Tyr  
1 5

<210> 12  
<211> 19  
<212> PRT  
<213> Peptides

<400> 12  
Phe Ile Trp Asn Lys Asp Ser Glu Tyr His Gly Gly Gly Asp Ala Trp  
1 5 10 15

Phe Gln Gly

<210> 13  
<211> 24  
<212> PRT  
<213> Peptides

<400> 13  
Asn Ala Phe Asp Phe Leu Leu Ala Asn Asp Val Asp Asn Ser Asn Pro  
1 5 10 15

Val Val Gln Ala Glu Asn Leu Asn  
20

<210> 14  
<211> 13  
<212> PRT  
<213> Peptides

<400> 14  
Ala Asn Phe Asp Ser Ile Arg Ile Asp Ala Val Asp Phe  
1 5 10

<210> 15  
<211> 8  
<212> PRT  
<213> Peptides

<400> 15  
His Ile Ser Leu Val Glu Ala Gly  
1 5

<210> 16  
<211> 8  
<212> PRT  
<213> Peptides

<400> 16  
Tyr Ser Ile Ile His Ala His Asp  
1 5

<210> 17  
<211> 8  
<212> PRT  
<213> Peptides

<400> 17  
Asp Val Val Asp Asn Gln Val Tyr  
1 5

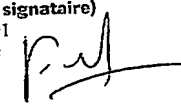
DÉPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08  
Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

DÉSIGNATION D'INVENTEUR(S) Page N° 1. / 2.  
(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

DA 113 W / 60899

Vos références pour ce dossier (facultatif)		B4787-FL	
N° D'ENREGISTREMENT NATIONAL		0103631	
<b>TITRE DE L'INVENTION</b> (200 caractères ou espaces maximum) MOLECULES D'ACIDES NUCLEIQUES CODANT UNE DEXTRANE- SACCHARASE CATALYSANT LA SYNTHÈSE DE DEXTRANE PORTANT DES RAMIFICATIONS DE TYPE ALPHA-1,2 OSIDIQUES.			
<b>LE(S) DEMANDEUR(S) :</b> INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE TOULOUSE CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE  représentés par :  ERNEST GUTMANN-YVES PLASSERAUD S.A. 3 rue Chauveau-Lagarde - 75008 PARIS			
<b>DESIGNE(NT) EN TANT QU'INVENTEUR(S) :</b> (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).			
Nom		MONSAN	
Prénoms		Pierre Emmanuel Frédéric	
Adresse	Rue	22 Chemin de la Gravette	
	Code postal et ville	31700	MONDONVILLE
Société d'appartenance (facultatif)			
Nom			
Prénoms			
Adresse	Rue		
	Code postal et ville		
Société d'appartenance (facultatif)			
Nom			
Prénoms			
Adresse	Rue		
	Code postal et ville		
Société d'appartenance (facultatif)			
<b>DATE ET SIGNATURE(S)</b> <b>DU (DES) DEMANDEUR(S)</b> <b>OU DU MANDATAIRE</b> (Nom et qualité du signataire) Paris, le 5 avril 2001 LAZARD Florence CPI n° 92.4029			

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire.  
Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.

DÉPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08

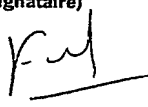
Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

DÉSIGNATION D'INVENTEUR(S) Page N° 1. / 2..

(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

DB 113 W / 260899

Vos références pour ce dossier (facultatif)		B4787-FL	
N° D'ENREGISTREMENT NATIONAL		0103631	
<b>TITRE DE L'INVENTION</b> (200 caractères ou espaces maximum) MOLECULES D'ACIDES NUCLEIQUES CODANT UNE DEXTRANE- SACCHARASE CATALYSANT LA SYNTHÈSE DE DEXTRANE PORTANT DES RAMIFICATIONS DE TYPE ALPHA-1,2 OSIDIQUES.			
<b>LE(S) DEMANDEUR(S) :</b> INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE TOULOUSE CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE  représentés par :  ERNEST GUTMANN-YVES PLASSERAUD S.A. 3 rue Chauveau-Lagarde - 75008 PARIS			
<b>DESIGNE(NT) EN TANT QU'INVENTEUR(S) :</b> (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).			
Nom		BOZONNET	
Prénoms		Sophie, Anne, Michèle	
Adresse	Rue	16 rue de la Gravette	
	Code postal et ville	31150	GAGNAC-SUR-GARONNE
Société d'appartenance (facultatif)			
Nom		REMAUD-SIMEON	
Prénoms		Magali, Martine, Claude	
Adresse	Rue	1 rue Benjamin Charrier	
	Code postal et ville	31520	RAMONVILLE SAINT-AGNE
Société d'appartenance (facultatif)			
Nom		WILLEMOT	
Prénoms		René-Marc, Lucien	
Adresse	Rue	3 Résidence Casteltrompette	
	Code postal et ville	31450	POMPERTUZAT
Société d'appartenance (facultatif)			
<b>DATE ET SIGNATURE(S)</b> <b>DU (DES) DEMANDEUR(S)</b> <b>OU DU MANDATAIRE</b> (Nom et qualité du signataire) Paris, le 5 avril 2001 LAZARD Florence CPI n° 92.4029			

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire.  
Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**